PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, C12P 19/32

A1

(11) International Publication Number: WO 98/42873

(43) International Publication Date: 1 October 1998 (01.10.98)

(21) International Application Number: PCT/US98/05809

(22) International Filing Date: 24 March 1998 (24.03.98)

(30) Priority Data:
08/823,516 24 March 1997 (24.03.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 08/823,516 (CIP)
Filed on 24 March 1997 (24.03.97)

(71) Applicant (*for all designated States except US*): THIRD WAVE TECHNOLOGIES, INC. [US/US]; 2800 S. Fish Hatchery Road, Madison, WI 53711 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): HALL, Jeff, G. [US/US]; 6305 Dylyn Drive, Madison, WI 53719 (US). LYAMICHEV, Victor I. [RU/US]; 2523 Carriedale Court, Madison, WI 53711 (US). MAST, Andrea, L. [US/US]; 2018 E. Washington Avenue, Madison, WI 53704 (US). BROW, Mary, Ann, D. [US/US]; 5905 Hammersley Road, Madison, WI 53711 (US). KWIATKOWSKI, Robert, W. [US/US]; 506 Azurene Lane, Verona, WI 53593 (US). VAVRA, Stephanie, H. [US/US]; 5601 Highway 12, Waunakee, WI 53597 (US).

(74) Agents: MACKNIGHT, Kamrin, T. et al.; Medlen & Carroll, LLP, Suite 2200, 220 Montgomery Street, San Francisco, CA 94104 (US).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: DETECTION OF NUCEIC ACIDS BY MULTIPLE SEQUENTIAL INVASIVE CLEAVAGES

(57) Abstract

The present invention relates to means for the detection and characterization of nucleic acid sequences, as well as variations in nucleic acid sequences. The present invention also relates to methods for forming a nucleic acid cleavage structure on a target sequence and cleaving the nucleic acid cleavage structure in a site-specific manner. The structure-specific nuclease activity of a variety of enzymes is used to cleave the target-dependent cleavage structure, thereby indicating the presence of specific nucleic acid sequences or specific variations thereof. The present invention further relates to methods and devices for the separation of nucleic acid molecules based on charge. The present invention also provides methods for the detection of non-target cleavage products via the formation of a complete and activated protein binding region. The invention further provides sensitive and specific methods for the detection of nucleic acid from various viruses in a sample.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

DETECTION OF NUCEIC ACIDS BY MULTIPLE SEQUENTIAL INVASIVE CLEAVAGES

FIELD OF THE INVENTION

The present invention relates to means for the detection and characterization of nucleic acid sequences and variations in nucleic acid sequences. The present invention relates to methods for forming a nucleic acid cleavage structure on a target sequence and cleaving the nucleic acid cleavage structure in a site-specific manner. The 5' nuclease activity of a variety of enzymes is used to cleave the target-dependent cleavage structure, thereby indicating the presence of specific nucleic acid sequences or specific variations thereof. The present invention further provides novel methods and devices for the separation of nucleic acid molecules based by charge. The present invention further provides methods for the detection of non-target cleavage products via the formation of a complete and activated protein binding region.

BACKGROUND OF THE INVENTION

The detection and characterization of specific nucleic acid sequences and sequence variations has been utilized to detect the presence of viral or bacterial nucleic acid sequences indicative of an infection, the presence of variants or alleles of mammalian genes associated with disease and cancers and the identification of the source of nucleic acids found in forensic samples, as well as in paternity determinations.

Various methods are known to the art which may be used to detect and characterize specific nucleic acid sequences and sequence variants. Nonetheless, as nucleic acid sequence data of the human genome, as well as the genomes of pathogenic organisms accumulates, the demand for fast, reliable, cost-effective and user-friendly tests for the detection of specific nucleic acid sequences continues to grow. Importantly, these tests must be able to create a detectable signal from samples which contain very few copies of the sequence of interest. The following discussion examines two levels of nucleic acid detection assays currently in use: I. Signal Amplification Technology for detection of rare sequences; and II. Direct Detection Technology for quantitative detection of sequences.

I. Signal Amplification Technology Methods For Amplification

The "Polymerase Chain Reaction" (PCR) comprises the first generation of methods for nucleic acid amplification. However, several other methods have been developed that employ the same basis of specificity, but create signal by different amplification mechanisms. These methods include the "Ligase Chain Reaction" (LCR), "Self-Sustained Synthetic Reaction" (3SR/NASBA), and "Qβ-Replicase" (Qβ).

Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR), as described in U.S. Patent Nos. 4,683,195 and 4,683,202 to Mullis and Mullis *et al.* (the disclosures of which are hereby incorporated by reference), describe a method for increasing the concentration of a segment of target sequence in a mixture of genomic DNA without cloning or purification. This technology provides one approach to the problems of low target sequence concentration. PCR can be used to directly increase the concentration of the target to an easily detectable level. This process for amplifying the target sequence involves introducing a molar excess of two oligonucleotide primers which are complementary to their respective strands of the double-stranded target sequence to the DNA mixture containing the desired target sequence. The mixture is denatured and then allowed to hybridize. Following hybridization, the primers are extended with polymerase so as to form complementary strands. The steps of denaturation, hybridization, and polymerase extension can be repeated as often as needed, in order to obtain relatively high concentrations of a segment of the desired target sequence.

The length of the segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and, therefore, this length is a controllable parameter. Because the desired segments of the target sequence become the dominant sequences (in terms of concentration) in the mixture, they are said to be "PCR-amplified."

Ligase Chain Reaction (LCR or LAR)

The ligase chain reaction (LCR; sometimes referred to as "Ligase Amplification Reaction" (LAR) described by Barany, Proc. Natl. Acad. Sci., 88:189 (1991); Barany, PCR Methods and Applic., 1:5 (1991); and Wu and Wallace, Genomics 4:560 (1989) has developed into a well-recognized alternative method for amplifying nucleic acids. In LCR, four oligonucleotides, two adjacent oligonucleotides which uniquely hybridize to one strand of target DNA, and a complementary set of adjacent oligonucleotides, which hybridize to the

opposite strand are mixed and DNA ligase is added to the mixture. Provided that there is complete complementarity at the junction, ligase will covalently link each set of hybridized molecules. Importantly, in LCR, two probes are ligated together only when they base-pair with sequences in the target sample, without gaps or mismatches. Repeated cycles of denaturation, hybridization and ligation amplify a short segment of DNA. LCR has also been used in combination with PCR to achieve enhanced detection of single-base changes. Segev, PCT Public. No. W09001069 A1 (1990). However, because the four oligonucleotides used in this assay can pair to form two short ligatable fragments, there is the potential for the generation of target-independent background signal. The use of LCR for mutant screening is limited to the examination of specific nucleic acid positions.

Self-Sustained Synthetic Reaction (3SR/NASBA)

The self-sustained sequence replication reaction (3SR) (Guatelli *et al.*, Proc. Natl. Acad. Sci., 87:1874-1878 [1990], with an erratum at Proc. Natl. Acad. Sci., 87:7797 [1990]) is a transcription-based *in vitro* amplification system (Kwok *et al.*, Proc. Natl. Acad. Sci., 86:1173-1177 [1989]) that can exponentially amplify RNA sequences at a uniform temperature. The amplified RNA can then be utilized for mutation detection (Fahy *et al.*, PCR Meth. Appl., 1:25-33 [1991]). In this method, an oligonucleotide primer is used to add a phage RNA polymerase promoter to the 5' end of the sequence of interest. In a cocktail of enzymes and substrates that includes a second primer, reverse transcriptase, RNase H, RNA polymerase and ribo-and deoxyribonucleoside triphosphates, the target sequence undergoes repeated rounds of transcription, cDNA synthesis and second-strand synthesis to amplify the area of interest. The use of 3SR to detect mutations is kinetically limited to screening small segments of DNA (*e.g.*, 200-300 base pairs).

Q-Beta (Qβ) Replicase

In this method, a probe which recognizes the sequence of interest is attached to the replicatable RNA template for Qβ replicase. A previously identified major problem with false positives resulting from the replication of unhybridized probes has been addressed through use of a sequence-specific ligation step. However, available thermostable DNA ligases are not effective on this RNA substrate, so the ligation must be performed by T4 DNA ligase at low temperatures (37°C). This prevents the use of high temperature as a

means of achieving specificity as in the LCR, the ligation event can be used to detect a mutation at the junction site, but not elsewhere.

Table 1 below, lists some of the features desirable for systems useful in sensitive nucleic acid diagnostics, and summarizes the abilities of each of the major amplification methods (*See also*, Landgren, Trends in Genetics 9:199 [1993]).

A successful diagnostic method must be very specific. A straight-forward method of controlling the specificity of nucleic acid hybridization is by controlling the temperature of the reaction. While the 3SR/NASBA, and Qβ systems are all able to generate a large quantity of signal, one or more of the enzymes involved in each cannot be used at high temperature (*i.e.*, >55°C). Therefore the reaction temperatures cannot be raised to prevent non-specific hybridization of the probes. If probes are shortened in order to make them melt more easily at low temperatures, the likelihood of having more than one perfect match in a complex genome increases. For these reasons, PCR and LCR currently dominate the research field in detection technologies.

TABLE 1

Feature	Method				
	PCR	LCR	PCR & LCR	3SR NASBA	Qβ
Amplifies Target	+	+	+	+	
Recognition of Independent Sequences Required	+	+	+	+	+
Performed at High Temp.	+	+			
Operates at Fixed Temp.				+	+
Exponential Amplification	+	+	+	+	+
Generic Signal Generation					+
Easily Automatable					

The basis of the amplification procedure in the PCR and LCR is the fact that the products of one cycle become usable templates in all subsequent cycles, consequently doubling the population with each cycle. The final yield of any such doubling system can be expressed as: $(1+X)^n = y$, where "X" is the mean efficiency (percent copied in each cycle), "n" is the number of cycles, and "y" is the overall efficiency, or yield of the reaction (Mullis, PCR Methods Applic., 1:1 [1991]). If every copy of a target DNA is utilized as a template in every cycle of a polymerase chain reaction, then the mean efficiency is 100%. If 20 cycles

of PCR are performed, then the yield will be 2^{20}, or 1,048,576 copies of the starting material. If the reaction conditions reduce the mean efficiency to 85%, then the yield in those 20 cycles will be only 1.85^{20}, or 220,513 copies of the starting material. In other words, a PCR running at 85% efficiency will yield only 21% as much final product, compared to a reaction running at 100% efficiency. A reaction that is reduced to 50% mean efficiency will yield less than 1% of the possible product.

In practice, routine polymerase chain reactions rarely achieve the theoretical maximum yield, and PCRs are usually run for more than 20 cycles to compensate for the lower yield. At 50% mean efficiency, it would take 34 cycles to achieve the million-fold amplification theoretically possible in 20, and at lower efficiencies, the number of cycles required becomes prohibitive. In addition, any background products that amplify with a better mean efficiency than the intended target will become the dominant products.

Also, many variables can influence the mean efficiency of PCR, including target DNA length and secondary structure, primer length and design, primer and dNTP concentrations, and buffer composition, to name but a few. Contamination of the reaction with exogenous DNA (*e.g.*, DNA spilled onto lab surfaces) or cross-contamination is also a major consideration. Reaction conditions must be carefully optimized for each different primer pair and target sequence, and the process can take days, even for an experienced investigator. The laboriousness of this process, including numerous technical considerations and other factors, presents a significant drawback to using PCR in the clinical setting. Indeed, PCR has yet to penetrate the clinical market in a significant way. The same concerns arise with LCR, as LCR must also be optimized to use different oligonucleotide sequences for each target sequence. In addition, both methods require expensive equipment, capable of precise temperature cycling.

Many applications of nucleic acid detection technologies, such as in studies of allelic variation, involve not only detection of a specific sequence in a complex background, but also the discrimination between sequences with few, or single, nucleotide differences. One method for the detection of allele-specific variants by PCR is based upon the fact that it is difficult for *Taq* polymerase to synthesize a DNA strand when there is a mismatch between the template strand and the 3' end of the primer. An allele-specific variant may be detected by the use of a primer that is perfectly matched with only one of the possible alleles; the mismatch to the other allele acts to prevent the extension of the primer, thereby preventing the amplification of that sequence. This method has a substantial limitation in that the base

composition of the mismatch influences the ability to prevent extension across the mismatch, and certain mismatches do not prevent extension or have only a minimal effect (Kwok *et al.*, Nucl. Acids Res., 18:999 [1990]).)

A similar 3'-mismatch strategy is used with greater effect to prevent ligation in the LCR (Barany, PCR Meth. Applic., 1:5 [1991]). Any mismatch effectively blocks the action of the thermostable ligase, but LCR still has the drawback of target-independent background ligation products initiating the amplification. Moreover, the combination of PCR with subsequent LCR to identify the nucleotides at individual positions is also a clearly cumbersome proposition for the clinical laboratory.

II. Direct Detection Technology

When a sufficient amount of a nucleic acid to be detected is available, there are advantages to detecting that sequence directly, instead of making more copies of that target, (*e.g.*, as in PCR and LCR). Most notably, a method that does not amplify the signal exponentially is more amenable to quantitative analysis. Even if the signal is enhanced by attaching multiple dyes to a single oligonucleotide, the correlation between the final signal intensity and amount of target is direct. Such a system has an additional advantage that the products of the reaction will not themselves promote further reaction, so contamination of lab surfaces by the products is not as much of a concern. Traditional methods of direct detection including Northern and Southern blotting and RNase protection assays usually require the use of radioactivity and are not amenable to automation. Recently devised techniques have sought to eliminate the use of radioactivity and/or improve the sensitivity in automatable formats. Two examples are the "Cycling Probe Reaction" (CPR), and "Branched DNA" (bDNA)

The cycling probe reaction (CPR) (Duck *et al.*, BioTech., 9:142 [1990]), uses a long chimeric oligonucleotide in which a central portion is made of RNA while the two termini are made of DNA. Hybridization of the probe to a target DNA and exposure to a thermostable RNase H causes the RNA portion to be digested. This destabilizes the remaining DNA portions of the duplex, releasing the remainder of the probe from the target DNA and allowing another probe molecule to repeat the process. The signal, in the form of cleaved probe molecules, accumulates at a linear rate. While the repeating process increases the signal, the RNA portion of the oligonucleotide is vulnerable to RNases that may be carried through sample preparation.

Branched DNA (bDNA), described by Urdea *et al.*, Gene 61:253-264 (1987), involves oligonucleotides with branched structures that allow each individual oligonucleotide to carry 35 to 40 labels (*e.g.*, alkaline phosphatase enzymes). While this enhances the signal from a hybridization event, signal from non-specific binding is similarly increased.

While both of these methods have the advantages of direct detection discussed above, neither the CPR or bDNA methods can make use of the specificity allowed by the requirement of independent recognition by two or more probe (oligonucleotide) sequences, as is common in the signal amplification methods described in Section I. above. The requirement that two oligonucleotides must hybridize to a target nucleic acid in order for a detectable signal to be generated confers an extra measure of stringency on any detection assay. Requiring two oligonucleotides to bind to a target nucleic acid reduces the chance that false "positive" results will be produced due to the non-specific binding of a probe to the target. The further requirement that the two oligonucleotides must bind in a specific orientation relative to the target,as is required in PCR, where oligonucleotides must be oppositely but appropriately oriented such that the DNA polymerase can bridge the gap between the two oligonucleotides in both directions, further enhances specificity of the detection reaction. However, it is well known to those in the art that even though PCR utilizes two oligonucleotide probes (termed primers) "non-specific" amplification (*i.e.*, amplification of sequences not directed by the two primers used) is a common artifact. This is in part because the DNA polymerase used in PCR can accommodate very large distances, measured in nucleotides, between the oligonucleotides and thus there is a large window in which non-specific binding of an oligonucleotide can lead to exponential amplification of inappropriate product. The LCR, in contrast, cannot proceed unless the oligonucleotides used are bound to the target adjacent to each other and so the full benefit of the dual oligonucleotide hybridization is realized.

An ideal direct detection method would combine the advantages of the direct detection assays (*e.g.*, easy quantification and minimal risk of carry-over contamination) with the specificity provided by a dual oligonucleotide hybridization assay.

SUMMARY OF THE INVENTION

The present invention relates to means for cleaving a nucleic acid cleavage structure in a site-specific manner. In a preferred embodiment, the means for cleaving is a structure-specific nuclease. Particularly preferred structure-specific nucleases are thermostable

structure-specific nucleases. In one embodiment, the structure-specific nuclease is an enzyme comprising 5' nucleases derived from thermostable DNA polymerases. These polymerases form the basis of a novel method of detection of specific nucleic acid sequences. The present invention contemplates use of novel detection methods for various uses, including, but not limited to clinical diagnostic purposes.

In one embodiment, the present invention contemplates a DNA sequence encoding a DNA polymerase altered in sequence (*i.e.*, a "mutant" DNA polymerase) relative to the native sequence, such that it exhibits altered DNA synthetic activity from that of the native (*i.e.*, "wild type") DNA polymerase. It is preferred that the encoded DNA polymerase is altered such that it exhibits reduced synthetic activity compared to that of the native DNA polymerase. In this manner, the enzymes of the invention are predominantly 5' nucleases and are capable of cleaving nucleic acids in a structure-specific manner in the absence of interfering synthetic activity.

Importantly, the 5' nucleases of the present invention are capable of cleaving linear duplex structures to create single discrete cleavage products. These linear structures are either 1) not cleaved by the wild type enzymes (to any significant degree), or 2) are cleaved by the wild type enzymes so as to create multiple products. This characteristic of the 5' nucleases has been found to be a consistent property of enzymes derived in this manner from thermostable polymerases across eubacterial thermophilic species.

It is not intended that the invention be limited by the nature of the alteration necessary to render the polymerase synthesis-deficient. Nor is it intended that the invention be limited by the extent of the deficiency. The present invention contemplates various structures, including altered structures (primary, secondary, etc.), as well as native structures, that may be inhibited by synthesis inhibitors.

Where the polymerase structure is altered, it is not intended that the invention be limited by the means by which the structure is altered. In one embodiment, the alteration of the native DNA sequence comprises a change in a single nucleotide. In another embodiment, the alteration of the native DNA sequence comprises a deletion of one or more nucleotides. In yet another embodiment, the alteration of the native DNA sequence comprises an insertion of one or more nucleotides. It is contemplated that the change in DNA sequence may manifest itself as change in amino acid sequence.

The present invention contemplates structure-specific nucleases from a variety of sources, including mesophilic, psychrophilic, thermophilic, and hyperthermophilic organisms.

The preferred structure-specific nucleases are thermostable. Thermostable structure-specific nucleases are contemplated as particularly useful in that they operate at temperatures where nucleic acid hybridization is extremely specific, allowing for allele-specific detection (including single-base mismatches). In one embodiment, the thermostable structure-specific are thermostable 5' nucleases which are selected from the group consisting of altered polymerases derived from the native polymerases of *Thermus* species, including, but not limited to *Thermus aquaticus*, *Thermus flavus,* and *Thermus thermophilus*. However, the invention is not limited to the use of thermostable 5' nucleases. Thermostable structure-specific nucleases from the FEN-1, RAD2 and XPG class of nucleases are also preferred.

Accordingly, the present invention provides improved enzymatic cleavage means. In one embodiment, the present invention provides a thermostable structure-specific nuclease having an amino acid sequence selected from the group consisting of SEQ ID NOS:61, 66, 69 and 72. In another embodiment, the nuclease is encoded by a DNA sequence selected from the group consisting of SEQ ID NO:60, 65, 68 and 70.

As noted above, the present invention contemplates the use of structure-specific nucleases in a detection method. In one embodiment, the present invention provides a method of detecting the presence of a target nucleic acid molecule comprising: a) providing: i) a cleavage means; ii) a source of a first target nucleic acid, the first target nucleic acid having a first region, a second region and a third region, wherein the first region is located adjacent to and downstream from the second region and wherein the second region is located adjacent to and downstream from the third region; iii) a first oligonucleotide having a 5' and a 3' portion wherein the 5' portion of the first oligonucleotide contains a sequence complementary to (at least a portion of) the second region of the first target nucleic acid and wherein the 3' portion of the first oligonucleotide contains a sequence complementary to at least a portion of the third region of the first target nucleic acid; iv) a second oligonucleotide having a 5' and a 3' portion wherein the 5' portion of the second oligonucleotide contains a sequence complementary to at least a portion of the first region of the first target nucleic acid and wherein the 3' portion of the second oligonucleotide contains a sequence complementary to at least a portion of the second region of the first target nucleic acid; v) a source of a second target nucleic acid, the second target nucleic acid having a first region, a second region and a third region, wherein the first region is located adjacent to and downstream from the second region and wherein the second region is located adjacent to and downstream from the third region; vi) a third oligonucleotide having a 5' and a 3' portion wherein the 5'

portion of the third oligonucleotide contains a sequence complementary to at least a portion of the second region of the second target nucleic acid and wherein the 3' portion of the third oligonucleotide contains a sequence complementary to at least a portion of the third region of the second target nucleic acid; b) generating a first cleavage structure wherein at least the 3' portion of the first oligonucleotide is annealed to the first target nucleic acid and wherein at least the 5' portion of the second oligonucleotide is annealed to the first target nucleic acid and wherein cleavage of the first cleavage structure occurs via the cleavage means thereby cleaving the first oligonucleotide to generate a fourth oligonucleotide, the fourth oligonucleotide having a 5' and a 3' portion wherein the 5' portion of the fourth oligonucleotide contains a sequence complementary to at least a portion of the first region of the second target nucleic acid and wherein the 3' portion of the fourth oligonucleotide contains a sequence complementary to at least a portion of the second region of the second target nucleic acid; c) generating a second cleavage structure under conditions wherein the at least the 3' portion of the third oligonucleotide is annealed to the second target nucleic acid and wherein at least the 5' portion of the fourth oligonucleotide is annealed to the second target nucleic acid oligonucleotide and wherein cleavage of the second cleavage structure occurs to generate a fifth oligonucleotide, the fifth oligonucleotide having a 3'-hydroxyl group; and d) detecting the fifth oligonucleotide.

It is contemplated that the first, second and third regions of the target nucleic acids be located adjacent to each other. However, the invention is not limited to the use of a target in which the three regions are contiguous with each other. Thus, the present invention contemplates the use of target nucleic acids wherein these three regions are contiguous with each other, as well as target acids wherein these three regions are not contiguous. It is further contemplated that gaps of approximately 2-10 nucleotides, representing regions of non-complementarity to the oligonucleotides (*e.g.*, the first and/or second oligonucleotides), may be present between the three regions of the target nucleic acid.

The methods of the present invention are not limited by the size of the oligonucleotides employed. In a preferred embodiment, the first oligonucleotide has a length between eleven and fifteen nucleotides.

It is intended that the generation of the first and second cleavage structures and cleavage of these structures occurs under a variety of conditions. In a preferred format, the conditions of generating the cleavage structures comprises mixing together the target nucleic acids with the first, second and third oligonucleotides and the cleavage means in an aqueous

solution in which a source of divalent cations is lacking. In this format, the cleavage reaction is initiated by the addition of a solution containing Mn^{2+} or Mg^{2+} ions. In another preferred format, the conditions of mixing comprises mixing together the target nucleic acid, and the first, second and third oligonucleotides in an aqueous solution containing Mn^{2+} or Mg^{2+} ions, and then adding the cleavage means to the reaction mixture.

It is contemplated that the oligonucleotides may be labelled. Thus, if the cleavage reaction employs a third oligonucleotide containing a label, detection of the cleavage product of the third oligonucleotide (*i.e.*, the fifth oligonucleotide) may comprise detection of the label. The invention is not limited by the nature of the label chosen, including, but not limited to, labels which comprise a dye or a radionucleotide (*e.g.*, ^{32}P), fluorescein moiety, a biotin moiety, luminogenic, fluorogenic, phosphorescent, or fluors in combination with moieties that can suppress emission by fluorescence energy transfer (FET). Numerous methods are available for the detection of nucleic acids containing any of the above-listed labels. For example, biotin-labeled oligonucleotide(s) may be detected using non-isotopic detection methods which employ streptavidin-alkaline phosphatase conjugates. Fluorescein-labelled oligonucleotide(s) may be detected using a fluorescein-imager. Further the oligonucleotide and particularly the probe oligonucleotides may contain positively charged adducts (*e.g.*, the Cy3 and Cy5 dyes, the dyes shown in Fig. 66, etc.) and/or positively charged amino acids and/or a phosphonate backbone to permit the detection of the fifth oligonucleotide (*i.e.*, the non-target cleavage product generated by cleavage of the second [or terminal if more than two reactions are employed in the cascade] cleavage structure) by selective charge reversal as described herein (*See,* section IV of the Description of the Invention). The oligonucleotides may be labelled with different labels (*e.g.*, the first and the third oligonucleotides may each bear a different label).

It is also contemplated that labelled oligonucleotides (cleaved or uncleaved) may be separated by means other than electrophoresis. For example, biotin-labelled oligonucleotides may be separated from nucleic acid present in the reaction mixture using para-magnetic or magnetic beads, or particles which are coated with avidin (or streptavidin). In this manner, the biotinylated oligonucleotide/avidin-magnetic bead complex can be physically separated from the other components in the mixture by exposing the complexes to a magnetic field. Additionally, the signal from the cleaved oligonucleotides may be resolved from that of the uncleaved oligonucleotides without physical separation. For example, a change in size, and

therefore rate of rotation in solution of fluorescent molecules can be detected by fluorescence polarization analysis.

In a preferred embodiment, the reaction conditions comprise a cleavage reaction temperature which is less than the melting temperature of the first oligonucleotide and greater than the melting temperature of the 3' portion of the first oligonucleotide. In a particularly preferred embodiment, the reaction temperature is between approximately 40-75°C; in another embodiment the reaction temperature is between approximately 40-60°C. It is contemplated that the reaction temperature at which the cleavage reaction occurs be selected with regard to the guidelines provided in the Description of the Invention.

The method of the present invention is not limited by the nature of the target nucleic acid. The target nucleic acid may comprise single-stranded or double-stranded DNA, RNA, and/or DNA/RNA hybrids. When a double-stranded target nucleic acid is employed, the reaction mixture may be treated such that the aid double-stranded DNA is rendered substantially single-stranded. A preferred method for rendering double-stranded DNA substantially single-stranded is by the use of increased temperature. When target nucleic acids comprising RNA are employed, the oligonucleotides may comprise DNA, RNA or an oligonucleotide comprising a mixture of RNA and DNA. It is not intended that the invention be limited by the nature of the oligonucleotides employed.

The oligonucleotides may comprise DNA, RNA or an oligonucleotide comprising a mixture of RNA and DNA. The invention also contemplates the use of a second oligonucleotide (*i.e.*, the upstream oligonucleotide in the first cleavage structure) which comprises a functional group (*e.g.*, a 5' peptide region) which prevents the dissociation of the 5' portion of the second oligonucleotide from the first region of the target nucleic acid. When such a functional group is present on the second oligonucleotide, the interaction between the 3' portion of the second oligonucleotide and the first region of the target nucleic acid may be destabilized (*i.e.*, designed to have a lower local melting temperature) through the use of A-T rich sequences, base analogs that form fewer hydrogen bonds (*e.g.*, dG-dU pairs) or through the use of phosphorothioate backbones, in order to allow the 5' region of the first oligonucleotide to compete successfully for hybridization.

The invention is not limited to use of oligonucleotides which are completely complementary to their cognate target sequences. In one embodiment, both the first and second oligonucleotides are completely complementary to the first target nucleic acid. In another embodiment, the first oligonucleotide is partially complementary to the first target

nucleic acid. In yet another embodiment, the second oligonucleotide is partially complementary to the first target nucleic acid. In yet another embodiment, both the first and the second oligonucleotide are partially complementary to the first target nucleic acid. Likewise, the third and fourth oligonucleotides may be either completely or partially complementary to the second target nucleic acid.

The methods of the invention may employ a source of target nucleic acid which comprises a sample containing genomic DNA. In a preferred embodiment, the sample containing genomic DNA is selected from the group including, but not limited to blood, saliva, cerebral spinal fluid, pleural fluid, milk, lymph, sputum and semen.

In a preferred embodiment, the method employs reaction conditions which comprise providing a source of divalent cations. In a particularly preferred embodiment, the divalent cation is selected from the group comprising Mn^{2+} and Mg^{2+} ions.

The invention is not limited by the nature of the cleavage means. As discussed above, the invention contemplates that the cleavage means comprises a thermostable 5' nuclease, although the invention is not limited to the use of a thermostable 5' nuclease. When a thermostable 5' nuclease is employed, a portion of the amino acid sequence of the nuclease may be homologous to a portion of the amino acid sequence of a thermostable DNA polymerase derived from a thermophilic organism. Particularly preferred cleavage means are structure-specific nucleases, with thermostable structure-specific nucleases being most preferred. In a preferred embodiment, the thermostable structure-specific nuclease is encoded by a DNA sequence selected from the group consisting of SEQ ID NOS:1-3, 9, 10, 12, 21, 25, 26, 60, 65, 68, 70, 74, and 78. In another preferred embodiment, the thermostable structure-specific nuclease is a nuclease from the FEN-1/RAD2/XPG class of nucleases. A preferred thermostable structure-specific nuclease is the *Pyrococcus woesii* FEN-1 endonuclease.

In another preferred embodiment, one or more of the first, second, and third oligonucleotides contain a dideoxynucleotide at the 3' terminus. When dideoxynucleotide-containing oligonucleotides are employed, the detection of the fifth oligonucleotide preferably comprises: a) incubating the fifth oligonucleotide with a template-independent polymerase and at least one labelled nucleoside triphosphate under conditions such that at least one labelled nucleotide is added to the 3'-hydroxyl group of the fifth oligonucleotide to generate a labelled fifth oligonucleotide; and b) detecting the presence of the labelled fifth oligonucleotide. The invention is not limited by the nature of the template-independent

polymerase employed; in one embodiment, the template-independent polymerase is selected from the group consisting of terminal deoxynucleotidyl transferase (TdT) and poly A polymerase. When TdT or polyA polymerase are employed in the detection step, the third oligonucleotide may contain a 5' end label, the 5' end label being a different label than the label present upon the labelled nucleoside triphosphate. The invention is not limited by the nature of the 5' end label; a wide variety of suitable 5' end labels are known to the art and include biotin, fluorescein, tetrachlorofluorescein, hexachlorofluorescein, Cy3, Cy5 and digoxigenin.

In another embodiment, detecting the fifth oligonucleotide comprises: a) incubating the fifth oligonucleotide with a template-independent polymerase and at least one nucleoside triphosphate under conditions such that at least one nucleotide is added to the 3'-hydroxyl group of the fifth oligonucleotide to generate a tailed fifth oligonucleotide; and b) detecting the presence of the tailed fifth oligonucleotide. The invention is not limited by the nature of the template-independent polymerase employed; in one embodiment, the template-independent polymerase is selected from the group consisting of terminal deoxynucleotidyl transferase (TdT) and poly A polymerase. When TdT or polyA polymerase are employed in the detection step, the second oligonucleotide may contain a 5' end label. The invention is not limited by the nature of the 5' end label; a wide variety of suitable 5' end labels are known to the art and include biotin, fluorescein, tetrachlorofluorescein, hexachlorofluorescein, Cy3, Cy5 and digoxigenin.

The invention further provides a method of detecting the fifth oligonucleotide (*i.e.*, the non-target cleavage product generated by cleavage of the second cleavage structure) comprising: a) providing: i) the fifth oligonucleotide; ii) a composition comprising two single-stranded nucleic acids annealed so as to define a single-stranded portion of a protein binding region; iii) a nucleic acid producing protein; b) exposing the fifth oligonucleotide to the single-stranded portion of the protein binding region under conditions such that the nucleic acid producing protein binds to the protein binding region and produces nucleic acid. In a preferred embodiment, the single-stranded portion of the protein binding region comprises: a) a first single continuous strand of nucleic acid comprising a sequence defining the template strand of an RNA polymerase binding region; and b) a second single continuous strand of nucleic acid having a 5' and a 3' end, the second nucleic acid comprising a region complementary to a portion of the first nucleic acid, wherein the second nucleic acid is

annealed to the first nucleic acid so as to define the single-stranded portion of the protein binding region.

The invention is not limited by the nature of the protein binding region employed. In a preferred embodiment, the protein binding region is a template-dependent RNA polymerase binding region, more preferably a T7 RNA polymerase binding region.

The invention further provides a method of detecting the fifth oligonucleotide comprising: a) providing: i) the fifth oligonucleotide; ii) a single continuous strand of nucleic acid comprising a sequence defining a single strand of an RNA polymerase binding region; iii) a template-dependent DNA polymerase; iv) a template-dependent RNA polymerase; b) exposing the fifth oligonucleotide to the RNA polymerase binding region under conditions such that the fifth oligonucleotide binds to a portion of the single strand of the RNA polymerase binding region; c) exposing the bound fifth oligonucleotide to the template-dependent DNA polymerase under conditions such that a double-stranded RNA polymerase binding region is produced; and d) exposing the double-stranded RNA polymerase binding region to the template-dependent RNA polymerase under conditions such that RNA transcripts are produced. In a preferred embodiment, the method further comprises detecting the RNA transcripts.

The invention is not limited by the nature of the protein binding region employed. In a preferred embodiment, the protein binding region is a template-dependent RNA polymerase binding region, more preferably a T7 RNA polymerase binding region.

The present invention also provides a method of detecting the presence of a target nucleic acid molecule comprising: a) providing: i) a cleavage means, ii) a source of a first target nucleic acid, the first target nucleic acid having a first region, a second region, a third region and a fourth region, wherein the first region is located adjacent to and downstream from the second region, the second region is located adjacent to and downstream from the third region and the third region is located adjacent to and downstream from the fourth region; iii) a first oligonucleotide complementary to (at least a portion of) the fourth region of the first target nucleic acid; iv) a second oligonucleotide having a 5' portion and a 3' portion wherein the 5' portion of the second oligonucleotide contains a sequence complementary to (at least a portion of) the second region of the first target nucleic acid and wherein the 3' portion of the second oligonucleotide contains a sequence complementary to (at least a portion of) the third region of the first target nucleic acid; iv) a third oligonucleotide having a 5' and a 3' portion wherein the 5' portion of the third oligonucleotide contains a sequence

complementary to (at least a portion of) the first region of the first target nucleic acid and wherein the 3' portion of the third oligonucleotide contains a sequence complementary to (at least a portion of) the second region of the first target nucleic acid; v) a source of a second target nucleic acid, the second target nucleic acid having a first region, a second region and a third region, wherein the first region is located adjacent to and downstream from the second region and wherein the second region is located adjacent to and downstream from the third region; vi) a fourth oligonucleotide having a 5' and a 3' portion wherein the 5' portion of the fourth oligonucleotide contains a sequence complementary to (at least a portion of) the second region of the second target nucleic acid and wherein the 3' portion of the fourth oligonucleotide contains a sequence complementary to (at least a portion of) the third region of the second target nucleic acid; b) generating a first cleavage structure wherein the first oligonucleotide is annealed to the fourth region of the first target nucleic acid and wherein at least the 3' portion of the second oligonucleotide is annealed to the first target nucleic acid and wherein at least the 5' portion of the third oligonucleotide is annealed to the first target nucleic acid and wherein cleavage of the first cleavage structure occurs thereby cleaving the second oligonucleotide to generate a fifth oligonucleotide, the fifth oligonucleotide having a 5' and a 3' portion wherein the 5' portion of the fifth oligonucleotide contains a sequence complementary to (at least a portion of) the first region of the second target nucleic acid and wherein the 3' portion of the fifth oligonucleotide contains a sequence complementary to (at least a portion of) the second region of the second target nucleic acid; c) generating a second cleavage structure under conditions wherein the at least the 3' portion of the fourth oligonucleotide is annealed to the second target nucleic acid and wherein at least the 5' portion of the fifth oligonucleotide is annealed to the second target nucleic acid and wherein cleavage of the second cleavage structure occurs to generate a sixth oligonucleotide, the sixth oligonucleotide having a 3'-hydroxyl group; and d) detecting the sixth oligonucleotide.

The detection of the sixth oligonucleotide may be accomplished by a variety of methods, such as those described above for the method in which the fifth oligonucleotide is to be detected. As described above, the invention is not limited by the nature of the target nucleic acids, the nature of the cleavage means, the nature of the oligonucleotides, etc.

The invention also provides a method of detecting the presence of a target nucleic acid molecule comprising: a) providing: i) a cleavage means, ii) a source of a target nucleic acid, the target nucleic acid having a first region, a second region and a third region, wherein the first region is located adjacent to and downstream from the second region and wherein the

second region is located adjacent to and downstream from the third region; iii) a first oligonucleotide having a 5' and a 3' portion wherein the 5' portion of the first oligonucleotide contains a sequence complementary to (at least a portion of) the second region of the target nucleic acid and wherein the 5' portion of the first oligonucleotide contains a region of self-complementarity and wherein the 3' portion of the first oligonucleotide contains a sequence complementary to (at least a portion of) the third region of the target nucleic acid; iv) a second oligonucleotide having a 5' and a 3' portion wherein the 5' portion of the second oligonucleotide contains a sequence complementary to (at least a portion of) the first region of the target nucleic acid and wherein the 3' portion of the second oligonucleotide contains a sequence complementary to (at least a portion of) the second region of the target nucleic acid; v) a third oligonucleotide having a 5' and a 3' portion wherein the 3' portion of the third oligonucleotide contains a sequence complementary to (at least a portion of) the 5' portion of the first oligonucleotide; b) generating a first cleavage structure wherein at least the 3' portion of the first oligonucleotide is annealed to the target nucleic acid and wherein at least the 5' portion of the second oligonucleotide is annealed to the target nucleic acid and wherein cleavage of the first cleavage structure occurs thereby cleaving the first oligonucleotide to generate a fourth oligonucleotide, the fourth oligonucleotide having a first region, a second region and a third region, wherein the first region is located adjacent to and upstream of the second region and wherein the second region is located adjacent to and upstream of the third region and wherein the third region of the fourth oligonucleotide contains a region of self-complementarity; c) generating a second cleavage structure under conditions wherein the at least the 3' portion of the third oligonucleotide is annealed to the first region of the fourth oligonucleotide and wherein at least the 5' portion of the fourth oligonucleotide is annealed to the second region of the third oligonucleotide and wherein the third region of the fourth oligonucleotide forms a hairpin structure and wherein cleavage of the second cleavage structure occurs to generate a fifth oligonucleotide, the fifth oligonucleotide having a 3'-hydroxyl group; and d) detecting the fifth oligonucleotide.

The detection of the fifth oligonucleotide may be accomplished by a variety of methods, such as those described above. As described above, the invention is not limited by the nature of the target nucleic acids, the nature of the cleavage means, the nature of the oligonucleotides, etc.

The invention further provides a method of detecting the presence of human cytomegalovirus (HCMV) nucleic acid in a sample comprising: a) providing: i) a cleavage means, ii) a sample suspected of containing human cytomegalovirus target nucleic acid, the target nucleic acid having a first region, a second region and a third region, wherein the first region is located adjacent to and downstream from the second region and wherein the second region is located adjacent to and downstream from the third region; iii) a first oligonucleotide having a 5' and a 3' portion wherein the 5' portion of the first oligonucleotide contains a sequence complementary to (at least a portion of) the second region of the target nucleic acid and wherein the 3' portion of the first oligonucleotide contains a sequence complementary to (at least a portion of) the third region of the target nucleic acid; iv) a second oligonucleotide having a 5' and a 3' portion wherein the 5' portion of the second oligonucleotide contains a sequence complementary to (at least a portion of) the first region of the target nucleic acid and wherein the 3' portion of the second oligonucleotide contains a sequence complementary to (at least a portion of) the second region of the target nucleic acid; b) generating a cleavage structure wherein at least the 3' portion of the first oligonucleotide is annealed to the target nucleic acid and wherein at least the 5' portion of the second oligonucleotide is annealed to the target nucleic acid and wherein cleavage of the cleavage structure occurs via the cleavage means to generate non-target cleavage products, each non-target cleavage product having a 3' hydroxyl group; and c) detecting the non-target cleavage products and thereby detecting the presence of human cytomegalovirus nucleic acid in the sample. In one embodiment, the first oligonucleotide has a length between eleven and fifteen nucleotides.

The detection of the fifth oligonucleotide (indicative of the presence of HCMV nucleic acid in the sample) may be accomplished by a variety of methods, such as those described above. As described above, the invention is not limited by the nature of the target nucleic acids, the nature of the cleavage means, the nature of the oligonucleotides, etc.

The invention also provides a method of detecting the presence of human cytomegalovirus nucleic acid in a sample comprising: a) providing: i) a cleavage means, ii) a sample suspected of containing human cytomegalovirus target nucleic acid, the target nucleic acid having a first region, a second region, a third region and a fourth region, wherein the first region is located adjacent to and downstream from the second region, the second region is located adjacent to and downstream from the third region and the third region is located adjacent to and downstream from the fourth region; iii) a first oligonucleotide complementary to (at least a portion of) the fourth region of the target nucleic acid; iv) a

second oligonucleotide having a 5' portion and a 3' portion wherein the 5' portion of the second oligonucleotide contains a sequence complementary to (at least a portion of) the second region of the target nucleic acid and wherein the 3' portion of the first oligonucleotide contains a sequence complementary to (at least a portion of) the third region of the target nucleic acid; v) a third oligonucleotide having a 5' portion and a 3' portion wherein the 5' portion of the third oligonucleotide contains a sequence complementary to (at least a portion of) the first region of the target nucleic acid and wherein the 3' portion of the third oligonucleotide contains a sequence complementary to (at least a portion of) the second region of the target nucleic acid; b) generating a cleavage structure wherein the first oligonucleotide is annealed to the fourth region of the target nucleic acid and wherein at least the 3' portion of the second oligonucleotide is annealed to the target nucleic acid and wherein at least the 5' portion of the third oligonucleotide is annealed to the target nucleic acid and wherein cleavage of the cleavage structure occurs via the cleavage means to generate non-target cleavage products, each non-target cleavage product having a 3' hydroxyl group; and c) detecting the non-target cleavage products and thereby detecting the presence of human cytomegalovirus nucleic acid in the sample. The detection of the non-target cleavage product (indicative of the presence of HCMV nucleic acid in the sample) may be accomplished by a variety of methods, such as those described above for the detection of the fifth oligonucleotide. As described above, the invention is not limited by the nature of the target nucleic acids, the nature of the cleavage means, the nature of the oligonucleotides, etc.

The present invention further provides a method of detecting the presence of a target nucleic acid molecule comprising: a) providing: a cleavage means; a source of a first target nucleic acid, wherein the first target nucleic acid has a first region, a second region and a third region, and wherein the first region is located adjacent to and downstream from the second region and wherein the second region is located adjacent to and downstream from the third region; first and second oligonucleotides having 3' and 5' portions, wherein the 3' portion of the first oligonucleotide contains a sequence complementary to the third region of the target nucleic acid and wherein the 5' portion of the first oligonucleotide and the 3' portion of the second oligonucleotide each contain sequence full complementary to the second region of the target nucleic acid, and wherein the 5' portion of the second oligonucleotide contains sequence complementary to the first region of the target nucleic acid; a source of a second target nucleic acid, the second target nucleic acid having a first region, a second region and a third region, wherein the first region is located adjacent to and downstream from

the second region and wherein the second region is located adjacent to and downstream from the third region; a third oligonucleotide having a 5' and a 3' portion wherein the 5' portion of the third oligonucleotide contains a sequence complementary to the second region of the second target nucleic acid and wherein the 3' portion of the third oligonucleotide contains a sequence complementary to the third region of the second target nucleic acid; b) generating a first cleavage structure wherein at least the 3' portion of the first oligonucleotide is annealed to the first target nucleic acid and wherein at least the 5' portion of the second oligonucleotide is annealed to the first target nucleic acid and wherein cleavage of the first cleavage structure occurs via the cleavage means thereby cleaving the first oligonucleotide to generate a fourth oligonucleotide, the fourth oligonucleotide having a 5' and a 3' portion wherein said 5' portion of the fourth oligonucleotide contains a sequence complementary to the first region of the second target nucleic acid and wherein the 3' portion of the fourth oligonucleotide contains a sequence complementary to the second region of the second target nucleic acid; c) generating a second cleavage structure under conditions wherein at least the 3' portion of the third oligonucleotide is annealed to the second target nucleic acid and wherein at least the 5' portion of the fourth oligonucleotide is annealed to the second target nucleic acid oligonucleotide and wherein cleavage of the second cleavage structure occurs to generate a fifth oligonucleotide, the fifth oligonucleotide having a 3'-hydroxyl group; and d) detecting the fifth oligonucleotide.

In some embodiments of this method, the first oligonucleotide has a length between eleven and fifteen nucleotides. In other embodiments, the cleavage means is a structure-specific nuclease. In yet other embodiments, the structure-specific nuclease is a thermostable structure-specific nuclease. In further embodiments, the thermostable structure-specific nuclease is an *Afu* FEN-1 endonuclease. In additional embodiments, the one or more first, second, and/or the third oligonucleotides contain a dideoxynucleotide at the 3' terminus. In particularly preferred embodiments, the method further comprises providing an Arrestor™, wherein the Arrestor™ reduces interaction between the first oligonucleotide and the second target.

In some preferred embodiments of the method the step of detecting the fifth oligonucleotide comprises: incubating the fifth oligonucleotide and at least one labelled nucleoside triphosphate under conditions such that at least one labelled nucleotide is added to the 3'-hydroxyl group of the fifth olignucleotide to generate a labelled fifth oligonucleotide; and detecting the presence of the labelled fifth oligonucleotide. In yet other preferred

embodiments of the method, the incubation is conducted in the presence of a polymerase. In alternative preferred embodiments, the polymerase is a template-dependent polymerase. In particularly preferred embodiments, the template-dependent polymerase is selected from the group consisting of terminal deoxynucleotidyl transferase and poly A polymerase. In yet other embodiments, the third oligonucleotide contains a 5' end label, the 5' end label being a different label than the label present upon the labelled nucleoside triphosphate.

In additional embodiments of the method the step of detecting the fifth oligonucleotide comprises: incubating the fifth oligonucleotide with a polymerase and at least one nucleoside triphosphate under conditions such that at least one nucleotide is added to the 3'-hydroxyl group of the fifth oligonucleotide to generate a tailed oligonucleotide; and detecting the presence of the tailed fifth oligonucleotide. In some preferred embodiments, the polymerase is a template-dependent polymerase. In other preferred embodiments, the template-dependent polymerase is selected from the group consisting of terminal deoxynucleotidyl transferase and poly A polymerase. In yet other embodiments, the third oligonucleotide contains a 5' end label.

In further embodiments of the method, the step of detecting the fifth oligonucleotide comprises: a) providing the fifth oligonucleotide, a single continuous strand of nucleic acid comprising a sequence definining a single strand of an RNA polymerase binding region, a template-dependent DNA polymerase, and a template-dependent RNA polymerase; b) exposing the fifth oligonucleotide to an RNA polymerase binding region under conditions such that the fifth oligonucleotide binds to a portion of the single strand of the RNA polymerase binding region to produce a bound fifth oligonucleotide; c) exposing the bound fifth oligonucleotide to the template-dependent DNA polymerase under conditions such that a double-stranded RNA polymerase binding region is produced; and d) exposing the double-stranded RNA polymerase binding region to the template-dependent RNA polymerase under conditions such that RNA transcripts are produced. In alternative embodiments of this method, the single-stranded portion of the protein binding region comprises: a first single continuous strand of nucleic acid comprising a sequence defining the template strand of an RNA polymerase binding region; and a second single continuous strand of nucleic acid having a 5' and a 3' end, the second nucleic acid comprising a region complementary to a portion of the first nucleic acid, wherein the second nucleic acid is annealed to the first nucleic acid so as to define the single-stranded portion of the protein binding region. In yet other embodiments of the method, the protein binding region is a template-dependent RNA

polymerase binding region. In still other embodiments, the template-dependent RNA polymerase binding region is the T7 RNA polymerase binding region.

In alternative preferred embodiments of the method, the step of detecting the fifth oligonucleotide comprises: providing a fifth oligonucleotide, a single continuous strand of nucleic acid comprising a sequence defining a single strand of an RNA polymerase binding region, a template-dependent DNA polymerase, and a template-dependent RNA polymerase; exposing the fifth oligonucleotide to an RNA polymerase binding region under conditions such that the fifth oligonucleotide binds to a portion of the single strand of the RNA polymerase binding region; exposing the bound fifth oligonucleotide to the template-dependent DNA polymerase under conditions such that a double-stranded RNA polymerase binding region is produced; and exposing the double-stranded RNA polymerase binding region to the template-dependent RNA polymerase under conditions such that RNA transcripts are produced. In preferred embodiments, the method further comprises detecting the RNA transcripts. In some preferred embodiments, the template-dependent RNA polymerase is T7 RNA polymerase.

The present invention also provides methods for detecting the presence of a target nucleic acid molecule. These methods comprise a) providing: a cleavage means; a source of a first target nucleic acid, wherein the first target nucleic acid has a first region, a second region, a third region and a fourth region, wherein the first region is located adjacent to and downstream from the second region, the second region is located adjacent to and downstream from the third region and the third region is located adjacent to and downstream from the fourth region; a first oligonucleotide complementary to the fourth region of the first target nucleic acid; second and third oligonucleotides having 3' and 5' portions, wherein the 3' portion of the second oligonucleotide contains a sequence complementary to the third region of the target nucleic acid and wherein the 5' portion of the second oligonucleotide and the 3' portion of the third oligonucleotide each contain sequence completely complementary to the second region of the target nucleic acid, and wherein the 5' portion of the third oligonucleotide contains sequence complementary to the first region of the target nucleic acid; a source of a second target nucleic acid, wherein the second target nucleic acid has a first region, a second region and a third region, wherein the first region is located adjacent to and downstream from the second region and wherein the second region is located adjacent to and downstream from the third region; a fourth oligonucleotide having a 5' and a 3' portion wherein the 5' portion of the fourth oligonucleotide contains a sequence complementary to

the second region of the second target nucleic acid and wherein the 3' portion of the fourth oligonucleotide contains a sequence complementary to the third region of the second target nucleic acid; b) generating a first cleavage structure wherein the first oligonucleotide is annealed to the fourth region of the first target nucleic acid and wherein at least the 3' portion of the second oligonucleotide is annealed to the first target nucleic acid and wherein at least the 5' portion of the third oligonucleotide is annealed to the first target nucleic acid and wherein cleavage of the first cleavage structure occurs thereby cleaving the second oligonucleotide to generate a fifth oligonucleotide, the fifth oligonucleotide having a 5' and a 3' portion wherein the 5' portion of the fifth oligonucleotide contains a sequence complementary to the first region of the second target nucleic acid and wherein the 3' portion of the fifth oligonucleotide contains a sequence complementary to the second region of the second target nucleic acid; c) generating a second cleavage structure under conditions wherein at least the 3' portion of the fourth oligonucleotide is annealed to the second target nucleic acid and wherein at least the 5' portion of the fifth oligonucleotide is annealed to the second target nucleic acid and wherein cleavage of the second cleavage structure occurs to generate a sixth oligonucleotide, the sixth oligonucleotide having a 3'-hydroxyl group; and d) detecting the sixth oligonucleotide. In some embodiments, the first oligonucleotide has a length between eleven and fifteen nucleotides. In yet other embodiments, the cleavage means is a structure-specific nuclease. In preferred embodiments, the structure-specific nuclease is a thermostable structure-specific nuclease. In particularly preferred embodiments, the thermostable structure-specific nuclease is an *Afu* FEN-1 endonuclease. In yet other embodiments, the one or more first, second, and/or third oligonucleotides contain a dideoxynucleotide at the 3' terminus. In particularly preferred embodiments, the method further comprises providing an Arrestor™, wherein the Arrestor™ reduces interaction between the first oligonucleotide and the second target.

In other embodiments of the method, the detecting of the sixth oligonucleotide comprises: incubating the sixth oligonucleotide and at least one labelled nucleoside triphosphate under conditions such that at least one labelled nucleotide is added to the 3'-hydroxyl group of the sixth olignucleotide to generate a labelled sixth oligonucleotide; and detecting the presence of the labelled sixth oligonucleotide. In preferred embodiments, the incubation step further comprises incubating a polymerase with the sixth oligonucleotide and at least one labelled nucleoside triphosphate. In particularly preferred embodiments, the polymerase is a template-dependent polymerase. In alternatively preferred embodiments, the

template-dependent polymerase is selected from the group consisting of terminal deoxynucleotidyl transferase and poly A polymerase.

DESCRIPTION OF THE DRAWINGS

Fig. 1 is a comparison of the nucleotide structure of the DNAP genes isolated from *Thermus aquaticus* (SEQ ID NO:1), *Thermus flavus* (SEQ ID NO:2) and *Thermus thermophilus* (SEQ ID NO:3); the consensus sequence (SEQ ID NO:7) is shown at the top of each row.

Fig. 2 is a comparison of the amino acid sequence of the DNAP isolated from *Thermus aquaticus* (SEQ ID NO:4), *Thermus flavus* (SEQ ID NO:5), and *Thermus thermophilus* (SEQ ID NO:6); the consensus sequence (SEQ ID NO:8) is shown at the top of each row.

Figs. 3A-G are a set of diagrams of wild-type and synthesis-deficient DNAP*Taq* genes.

Fig. 4A depicts the wild-type *Thermus flavus* polymerase gene.

Fig. 4B depicts a synthesis-deficient *Thermus flavus* polymerase gene.

Fig. 5 depicts a structure which cannot be amplified using DNAP*Taq*; this Figure shows SEQ ID NO:17 (primer) and SEQ ID NO:15 (hairpin).

Fig. 6 is a ethidium bromide-stained gel demonstrating attempts to amplify a bifurcated duplex using either DNAP*Taq* or DNAPStf (*i.e.*, the Stoffel fragment of DNAP*Taq*).

Fig. 7 is an autoradiogram of a gel analyzing the cleavage of a bifurcated duplex by DNAP*Taq* and lack of cleavage by DNAPStf.

Figs. 8A-B are a set of autoradiograms of gels analyzing cleavage or lack of cleavage upon addition of different reaction components and change of incubation temperature during attempts to cleave a bifurcated duplex with DNAP*Taq*.

Figs. 9A-B are an autoradiogram displaying timed cleavage reactions, with and without primer.

Figs. 10A-B are a set of autoradiograms of gels demonstrating attempts to cleave a bifurcated duplex (with and without primer) with various DNAPs.

Fig. 11A shows the substrate and oligonucleotides (19-12 [SEQ ID NO:18] and 30-12 [SEQ ID NO:19]) used to test the specific cleavage of substrate DNAs targeted by pilot oligonucleotides.

Fig. 11B shows an autoradiogram of a gel showing the results of cleavage reactions using the substrates and oligonucleotides shown Fig. 12A.

Fig. 12A shows the substrate and oligonucleotide (30-0 [SEQ ID NO:20]) used to test the specific cleavage of a substrate RNA targeted by a pilot oligonucleotide.

Fig. 12B shows an autoradiogram of a gel showing the results of a cleavage reaction using the substrate and oligonucleotide shown in Fig. 13A.

Fig. 13 is a diagram of vector pTTQ18.

Fig. 14 is a diagram of vector pET-3c.

Figs. 15A-E depicts a set of molecules which are suitable substrates for cleavage by the 5' nuclease activity of DNAPs (SEQ ID NOS:15 and 17 are depicted in Fig.15E).

Fig. 16 is an autoradiogram of a gel showing the results of a cleavage reaction run with synthesis-deficient DNAPs.

Fig. 17 is an autoradiogram of a PEI chromatogram resolving the products of an assay for synthetic activity in synthesis-deficient DNAP*Taq* clones.

Fig. 18A depicts the substrate molecule (SEQ ID NOS:15 and 17) used to test the ability of synthesis-deficient DNAPs to cleave short hairpin structures.

Fig. 18B shows an autoradiogram of a gel resolving the products of a cleavage reaction run using the substrate shown in Fig. 19A.

Fig. 19 provides the complete 206-mer duplex sequence (SEQ ID NO:27) employed as a substrate for the 5' nucleases of the present invention

Figs. 20A and B show the cleavage of linear nucleic acid substrates (based on the 206-mer of Fig. 21) by wild type DNAPs and 5' nucleases isolated from *Thermus aquaticus* and *Thermus flavus*.

Fig. 21A shows the "nibbling" phenomenon detected with the DNAPs of the present invention.

Fig. 21B shows that the "nibbling" of Fig. 25A is 5' nucleolytic cleavage and not phosphatase cleavage.

Fig. 22 demonstrates that the "nibbling" phenomenon is duplex dependent.

Fig. 23 is a schematic showing how "nibbling" can be employed in a detection assay.

Figs. 24A and B demonstrates that "nibbling" can be target directed.

Fig. 25 provides a schematic drawing of a target nucleic acid with an Invader™ oligonucleotide and a probe oligonucleotide annealed to the target.

Fig. 26 provides a schematic showing the S-60 hairpin oligonucleotide (SEQ ID NO:29) with the annealed P-15 oligonucleotide (SEQ ID NO:30).

Fig. 27 is an autoradiogram of a gel showing the results of a cleavage reaction run using the S-60 hairpin in the presence or absence of the P-15 oligonucleotide.

Fig. 28 provides a schematic showing three different arrangements of target-specific oligonucleotides and their hybridization to a target nucleic acid which also has a probe oligonucleotide annealed thereto (SEQ ID NOS:31-35).

Fig. 29 is the image generated by a fluoroscence imager showing that the presence of an Invader™ oligonucleotide causes a shift in the site of cleavage in a probe/target duplex.

Fig. 30 is the image generated by a fluoroscence imager showing the products of Invader™-directed cleavage assays run using the three target-specific oligonucleotides diagrammed in Fig. 28.

Fig. 31 is the image generated by a fluoroscence imager showing the products of Invader™-directed cleavage assays run in the presence or absence of non-target nucleic acid molecules.

Fig. 32 is the image generated by a fluoroscence imager showing the products of Invader™-directed cleavage assays run in the presence of decreasing amounts of target nucleic acid.

Fig. 33 is the image generated by a fluoroscence imager showing the products of Invader™-directed cleavage assays run in the presence or absence of saliva extract using various thermostable 5' nucleases or DNA polymerases.

Fig. 34 is the image generated by a fluoroscence imager showing the products of Invader™-directed cleavage assays run using various 5' nucleases.

Fig. 35 is the image generated by a fluoroscence imager showing the products of Invader™-directed cleavage assays run using two target nucleic acids which differ by a single basepair at two different reaction temperatures.

Fig. 36A provides a schematic showing the effect of elevated temperature upon the annealing and cleavage of a probe oligonucleotide along a target nucleic acid wherein the probe contains a region of noncomplementarity with the target.

Fig. 36B provides a schematic showing the effect of adding an upstream oligonucleotide upon the annealing and cleavage of a probe oligonucleotide along a target nucleic acid wherein the probe contains a region of noncomplementarity with the target.

Fig. 37 provides a schematic showing an arrangement of a target-specific Invader™ oligonucleotide (SEQ ID NO:39) and a target-specific probe oligonucleotide (SEQ ID NO:38) bearing a 5' Cy3 label along a target nucleic acid (SEQ ID NO:31).

Fig. 38 is the image generated by a fluorescence imager showing the products of Invader™-directed cleavage assays run in the presence of increasing concentrations of KCl.

Fig. 39 is the image generated by a fluorescence imager showing the products of Invader™-directed cleavage assays run in the presence of increasing concentrations of $MnCl_2$ or $MgCl_2$.

Fig. 40 is the image generated by a fluorescence imager showing the products of Invader™-directed cleavage assays run in the presence of increasing amounts of genomic DNA or tRNA.

Fig. 41 is the image generated by a fluorescence imager showing the products of Invader™-directed cleavage assays run use a HCV RNA target.

Fig. 42 is the image generated by a fluorescence imager showing the products of Invader™-directed cleavage assays run using a HCV RNA target and demonstrate the stability of RNA targets under Invader™-directed cleavage assay conditions.

Fig. 43 is the image generated by a fluorescence imager showing the sensitivity of detection and the stability of RNA in Invader™-directed cleavage assays run using a HCV RNA target.

Fig. 44 is the image generated by a fluorescence imager showing thermal degradation of oligonucleotides containing or lacking a 3' phosphate group.

Fig. 45 depicts the structure of amino-modified oligonucleotides 70 and 74.

Fig. 46 depicts the structure of amino-modified oligonucleotide 75

Fig. 47 depicts the structure of amino-modified oligonucleotide 76.

Fig. 48 is the image generated by a fluorescence imager scan of an IEF gel showing the migration of substrates 70, 70dp, 74, 74dp, 75, 75dp, 76 and 76dp.

Fig. 49A provides a schematic showing an arrangement of a target-specific Invader™ oligonucleotide (SEQ ID NO:50) and a target-specific probe oligonucleotide (SEQ ID NO:51) bearing a 5' Cy3 label along a target nucleic acid (SEQ ID NO:52).

Fig. 49B is the image generated by a fluorescence imager showing the detection of specific cleavage products generated in an invasive cleavage assay using charge reversal (*i.e.*, charge based separation of cleavage products).

Fig. 50 is the image generated by a fluorescence imager which depicts the sensitivity of detection of specific cleavage products generated in an invasive cleavage assay using charge reversal.

Fig. 51 depicts a first embodiment of a device for the charge-based separation of oligonucleotides.

Fig. 52 depicts a second embodiment of a device for the charge-based separation of oligonucleotides.

Fig. 53 shows an autoradiogram of a gel showing the results of cleavage reactions run in the presence or absence of a primer oligonucleotide; a sequencing ladder is shown as a size marker.

Figs. 54A-D depict four pairs of oligonucleotides; in each pair shown, the upper arrangement of a probe annealed to a target nucleic acid lacks an upstream oligonucleotide and the lower arrangement contains an upstream oligonucleotide (SEQ ID NOS:32 and 54-58 are shown in Figs. 54A-D).

Fig. 55 shows the chemical structure of several positively charged heterodimeric DNA-binding dyes.

Fig. 56 is a schematic showing alternative methods for the tailing and detection of specific cleavage products in the context of the Invader™-directed cleavage assay.

Fig. 57 provides a schematic drawing of a target nucleic acid with an Invader™ oligonucleotide, a miniprobe, and a stacker oligonucleotide annealed to the target.

Fig. 58 provides a space-filling model of the 3-dimensional structure of the T5 5'-exonuclease.

Fig. 59 provides an alignment of the amino acid sequences of several FEN-1 nucleases including the *Methanococcus jannaschii* FEN-1 protein (MJAFEN1.PRO), the *Pyrococcus furiosus* FEN-1 protein (PFUFEN1.PRO), the human FEN-1 protein (HUMFEN1.PRO), the mouse FEN-1 protein (MUSFEN1.PRO), the *Saccharomyces cerevisiae* YKL510 protein (YST510.PRO), the *Saccharomyces cerevisiae* RAD2 protein (YSTRAD2.PRO), the *Shizosaccharomyces pombe* RAD13 protein (SPORAD13.PRO), the human XPG protein (HUMXPG.PRO), the mouse XPG protein (MUSXPG.PRO), the *Xenopus laevis* XPG protein (XENXPG.PRO) and the *C. elegans* RAD2 protein (CELRAD2.PRO) (SEQ ID NOS:135-145, respectively); portions of the amino acid sequence of some of these proteins were not shown in order to maximize the alignment between proteins (specifically, amino acids 122 to 765 of the YSTRAD2 sequence were deleted; amino acids 122 to 746 of the SPORAD13

sequence were deleted; amino acids 122 to 757 of the HUMXPG sequence were deleted; amino acids 122 to 770 of the MUSXPG sequence were deleted; and amino acids 122 to 790 of the XENXPG sequence were deleted). The numbers to the left of each line of sequence refers to the amino acid residue number; dashes represent gaps introduced to maximize alignment.

Fig. 60 is a schematic showing the S-33 (SEQ ID NO:84) and 11-8-0 (SEQ ID NO:85) oligonucleotides in a folded configuration; the cleavage site is indicated by the arrowhead.

Fig. 61 shows a Coomassie stained SDS-PAGE gel showing the thrombin digestion of Cleavase® BN/thrombin.

Fig. 62 is the image generated by a fluorescence imager showing the products produced by the cleavage of the S-60 hairpin using Cleavase® BN/thrombin (before and after thrombin digestion).

Fig. 63 is the image generated by a fluorescence imager showing the products produced by the cleavage of circular M13 DNA using Cleavase® BN/thrombin.

Fig. 64 is an SDS-PAGE gel showing the migration of purified Cleavase® BN nuclease, *Pfu* FEN-1, *Pwo* FEN-1 and *Mja* FEN-1.

Fig. 65 is the image generated by a fluorescence imager showing the products produced by the cleavage of the S-33 and 11-8-0 oligonucleotides by Cleavase® BN and the *Mja* FEN-1 nucleases.

Fig. 66 is the image generated by a fluorescence imager showing the products produced by the incubation of an oligonucleotide either having or lacking a 3'-OH group with TdT.

Fig. 67 is the image generated by a fluorescence imager showing the products produced the incubation of cleavage products with TdT.

Fig. 68 is a photograph of a Universal GeneComb™ card showing the capture and detection of cleavage products on a nitrocellulose support.

Fig. 69 is the image generated by a fluorescence imager showing the products produced using the Cleavase® A/G and *Pfu* FEN-1 nucleases and a fluorescein-labeled probe.

Fig. 70 is the image generated by a fluorescence imager showing the products produced using the Cleavase® A/G and *Pfu* FEN-1 nucleases and a Cy3-labeled probe.

Fig. 71 is the image generated by a fluorescence imager showing the products produced using the Cleavase® A/G and *Pfu* FEN-1 nucleases and a TET-labeled probe.

Figs. 72A and 72B are images generated by a fluorescence imager showing the products produced using the Cleavase® A/G and *Pfu* FEN-1 nucleases and probes having or lacking a 5' positive charge; the gel shown in Fig. 83A was run in the standard direction and the gel shown in Fig. 84B was run in the reverse direction.

Fig. 73 shows the structure of 3-nitropyrrole and 5-nitroindole.

Fig. 74 shows the sequence of oligos 109, 61 and 67 (SEQ ID NOS:97, 50 and 51) annealed into a cleavage structure as well as the sequence of oligo 67 (SEQ ID NO:51) and a composite of SEQ ID NOS:98, 99, 101 and 102.

Fig. 75A-C show images generated by a fluorescence imager showing the products produced in an Invader™-directed cleavage assay performed at various temperatures using a miniprobe which is either completely complementary to the target or contains a single mismatch with the target.

Fig. 76 shows the sequence of oligos 166 (SEQ ID NO:103), 165 (SEQ ID NO:104), 161 (SEQ ID NO:106), 162 (SEQ ID NO:105) and 164 (SEQ ID NO:107) as well as a cleavage structure.

Fig. 77 shows the image generated by a fluorescence imager showing the products produced in an Invader™-directed cleavage assay performed using *ras* gene sequences as the target.

Figs. 78A-C show the sequence of the S-60 hairpin (SEQ ID NO:29) (A), and the P-15 oligo (SEQ ID NO:30) (shown annealed to the S-60 hairpin in B) and the image generated by a fluorescence imager showing the products produced by cleavage of the S-60 hairpin in the presence of various Invader™ oligos.

Fig. 79 shows the structure of various 3' end substituents.

Fig. 80 is a composite graph showing the effect of probe concentration, temperature and a stacker oligonucleotide on the cleavage of miniprobes.

Fig. 81 shows the sequence of the IT-2 oligonucleotide (SEQ ID NO:115; shown in a folded configuration) as well as the sequence of the IT-1 (SEQ ID NO:116) and IT-A (SEQ ID NO:117) oligos.

Fig. 82 shows the image generated by a fluorescence imager showing the products produced by cleavage of the oligos shown in Fig. 92 by Cleavase® A/G nuclease.

Fig. 83 shows the image generated by a fluorescence imager which provides a comparison of the rates of cleavage by the *Pfu* FEN-1 and *Mja* FEN-1 nucleases.

Fig. 84 shows the image generated by a fluorescence imager which depicts the detection of RNA targets using a miniprobe and stacker oligonucleotides.

Figs. 85A-C provide schematics showing particular embodiments of the present invention wherein a T7 promoter region and copy template annealed with either no oligo (A), a complete promoter oligo (B) or a complete promoter oligo with a 3' tail (C); one strand of the T7 promoter region is indicated by the hatched line.

Figs. 86A-D provide schematics showing particular embodiments of the present invention wherein a T7 promoter region and copy template annealed with either a cut probe(A), a partial promoter oligo (B), an uncut oligo (C) or both an uncut probe and a partial promoter oligo (D).

Fig. 87 provides a schematic illustrating one embodiment of the present invention wherein a template-dependent DNA polymerase is used to extend a cut probe to complete a T7 promoter region and thereby allow transcription.

Fig. 88 provides a schematic illustrating that an uncut probe combined with a partial promoter oligo does not permit transcription while a cut probe combined with a partial promoter oligo generates a complete (but nicked) promoter which supports transcription.

Fig. 89 shows the image generated by a fluorescence imager which shows that primer extension can be used to complete a partial promoter formed by a cut probe (lanes 1-5) and that annealing a cut probe generated in an invasive cleavage assay can complete a partial T7 promoter to permit transcription (lanes 6-9).

Figs. 90A-C provide schematics showing particular embodiments of the present invention which illustrate that the use of a partial promoter oligo with a paired 5' tail can be used to block transcription from a composite promoter formed by the annealing of an uncut probe.

Fig. 91 shows the image generated by a fluorescence imager which shows that transcription from a "leaky" branched T7 composite promoter can be shut down by the use of a downstream partial promoter oligo having a paired 5' tail.

Fig. 92 shows the image generated by a fluorescence imager which shows that the location of the nick site in a nicked composite T7 promoter can effect the efficiency of transcription.

Fig. 93 shows the image generated by a fluorescence imager which shows that the presence of an unpaired 3' tail on a full-length promoter oligo decreases but does not abolish

transcription. Beneath the image are schematics showing the nucleic acids tested in reactions 1-4; these schematics show SEQ ID NOS:123-125.

Fig. 94 is a schematic which illustrates one embodiment of the present invention where a composite T7 promoter region is created by the binding of the cut probe oligo downstream of the partial promoter oligo.

Figs. 95A-D provide schematics showing particular embodiments of the present invention which show various ways in which a composite promoter can be formed wherein the nick is located in the template (or bottom) strand.

Fig. 96 is a schematic which illustrates one embodiment of the present invention where the cut probe from an initial invasive cleavage reaction is employed as the Invader™ oligonucleotide in a second invasive cleavage reaction.

Fig. 97 is a schematic which illustrates one embodiment of the present invention where the cut probe from an initial invasive cleavage reaction is employed as an integrated Invader™-target complex in a second invasive cleavage reaction.

Fig. 98 shows the nucleotide sequence of the PR1 probe (SEQ ID NO:119), the IT3 Invader™-Target oligonculeotide (SEQ ID NO:118), the IT3-8, IT3-6, IT3-4, IT3-3 and IT3-0 oligonucleotides (SEQ ID NOS:147-151, respectively).

Fig. 99 depicts structures that may be employed to determine the ablity of an enzyme to cleave a probe in the presence and the absence of an upstream oligonucleotide. Fig. 99 displays the sequence of oligo 89-15-1 (SEQ ID NO:152), oligo 81-69-5 (SEQ ID NO:156), oligo 81-69-4 (SEQ ID NO155), oligo 81-69-3 (SEQ ID NO:154), oligo 81-69-2 (SEQ ID NO:153) and a portion of M13mp18 (SEQ ID NO:163).

Fig. 100 shows the image generated by a fluorescence imager which shows the dependence of *Pfu* FEN-1 on the presence of an overlapping upstream oligonucleotide for specific cleavage of the probe.

Fig. 101a shows the image generated by a fluorescence imager which compares the amount of product generated in a standard (*i.e.*, a non-sequential invasive cleavage reaction) and a sequential invasive cleavage reaction.

Figure 101b is a graph comparing the amount of product generated in a standard or basic (*i.e.*, a non-sequential invasive cleavage reaction) and a sequential invasive cleavage reaction ("invader sqrd") (y axis = fluorescence units; x axix = attomoles of target).

Fig. 102 shows the image generated by a fluorescence imager which shows that the products of a completed sequential invasive cleavage reaction cannot cross contaminant a subsequent similar reaction.

Fig. 103 shows the sequence of the oligonucleotide employed in an invasive cleavage reaction for the detection of HCMV viral DNA; Fig. 103 shows the sequence of oligo 89-76 (SEQ ID NO:161), oligo 89-44 (SEQ ID NO:160) and nucleotides 3057-3110 of the HCMV genome (SEQ ID NO:162).

Fig. 104 shows the image generated by a fluorescence imager which shows the sensitive detection of HCMV viral DNA in samples containing human genomic DNA using an invasive cleavage reaction.

Fig. 105 is a schematic which illustrates one embodiment of the present invention, where the cut probe from an initial invasive cleavage reaction is employed as the Invader™ oligonucleotide in a second invasive cleavage reaction, and where an Arrestor™ oligonucleotide prevents participation of remaining uncut first probe in the cleavage of the second probe.

Fig. 106 is a schematic which illustrates one embodiment of the present invention, where the cut probe from an initial invasive cleavage reaction is employed as an integrated Invader™-target complex in a second invasive cleavage reaction, and where an Arrestor™ oligonucleotide prevents participation of remaining uncut first probe in the cleavage of the second probe.

Fig. 107 shows three images generated by a fluorescence imager showing that two different lengths of 2' O-methyl, 3' terminal amine-modified Arrestor™ oligonucleotide both reduce non-specific background cleavage of the secondary probe when included in the second step of a reaction where the cut probe from an initial invasive cleavage reaction is employed as an integrated Invader™-target complex in a second invasive cleavage reaction.

Fig. 108A shows two images generated by a fluorescence imager showing the effects on nonspecific and specific cleavage signal of increasing concentrations of primary probe in the first step of a reaction where the cut probe from an initial invasive cleavage reaction is employed as the Invader™ oligonucleotide in a second invasive cleavage reaction.

Fig. 108B shows two images generated by a fluorescence imager showing the effects on nonspecific and specific cleavage signal of increasing concentrations of primary probe in the first step of a reaction, and inclusion of a 2' O-methyl, 3' terminal amine-modified Arrestor™ oligonucleotide in the second step of a reaction where the cut probe from an initial

invasive cleavage reaction is employed as the Invader™ oligonucleotide in a second invasive cleavage reaction.

Fig. 108C shows shows a graph generated using the spreadsheet Microsoft Excel software, comparing the effects on nonspecific and specific cleavage signal of increasing concentrations of primary probe in the first step of a reaction , in the presence or absence of a 2' O-methyl, 3' terminal amine-modified Arrestor™ oligonucleotide in the second step of a reaction where the cut probe from an initial invasive cleavage reaction is employed as the Invader™ oligonucleotide in a second invasive cleavage reaction.

Fig. 109A shows two images generated by a fluorescence imager showing the effects on nonspecific and specific cleavage signal of including an unmodified Arrestor™ oligonucleotide in the second step of a reaction where the cut probe from an initial invasive cleavage reaction is employed as the Invader™ oligonucleotide in a second invasive cleavage reaction.

Fig. 109B shows two images generated by a fluorescence imager showing the effects on nonspecific and specific cleavage signal of including a 3' terminal amine modified Arrestor™, a partially 2' O-methyl substituted, 3' terminal amine modified Arrestor™, or an entirely 2' O-methyl, 3' terminal amine modified Arrestor™ oligonucleotide in the second step of a reaction where the cut probe from an initial invasive cleavage reaction is employed as the Invader™ oligonucleotide in a second invasive cleavage reaction.

Fig. 110A shows two images generated by a fluorescence imager comparing the effects on nonspecific and specific cleavage signal of including an Arrestor™ oligonucleotides of different lengths in the second step of a reaction where the cut probe from an initial invasive cleavage reaction is employed as the Invader™ oligonucleotide in a second invasive cleavage reaction.

Figure 110B shows two images generated by a fluorescence imager comparing the effects on nonspecific and specific cleavage signal of including an arrestoer oligonucleotides of different lengths in the second step of a reaction where the cut probe from an initial invasive cleavage reaction is employed as the Invader™ oligonucleotide in a second invasive cleavage reaction, and in which a longer variant of the secondary probe used in the reactions in Fig. 110A is tested.

Fig. 110C shows a schematic diagram of a primary probe aligned with several Arrestor™ oligonucleotides of different lengths. The region of the primary probe that is

complementary to the HBV target sequence is underlined. The Arrestors™ are aligned with the probe by complementarity.

Fig. 111 shows two images generated by a fluorescence imager comparing the effects on nonspecific and specific cleavage signal of including Arrestor™ oligonucleotides of different lengths in the second step of a reaction where the cut probe from an initial invasive cleavage reaction is employed as the Invader™ oligonucleotide in a second invasive cleavage reaction, using secondary probes of two different lengths.

DEFINITIONS

As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (*i.e.*, a sequence of nucleotides such as an oligonucleotide or a target nucleic acid) related by the base-pairing rules. For example, for the sequence "A-G-T," is complementary to the sequence "T-C-A." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods which depend upon binding between nucleic acids.

The term "self-complementarity" when used in reference to a nucleic acid strand (*e.g.*, an oligonucleotide) means that separate regions of that strand can base-pair. Because this term refers only to intramolecular base-pairing, any strand said to have a region of self-complementarity must have at least two regions capable of base-pairing with one another. As defined above, complementarity may be either "complete" or "partial". As used in reference to the probe oligonucleotides of the present invention, regions are considered to have significant self-complementarity when they may form a duplex of at least 3 contiguous base pairs (*i.e.*, three base pairs of complete complementarity), or when they may form a longer duplex that is partially complementary. The ability of an oligonucleotide having a region of self-complementarity to successfully serve both as a target strand for a probe, and as an upstream oligonucleotide that directs invasive cleavage of that probe as depicted in Fig. 97, is considered sufficient demonstration of self-complementarity as defined herein.

The term "homology" refers to a degree of identity. There may be partial homology or complete homology. A partially identical sequence is one that is less than 100% identical to another sequence.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids.

As used herein, the term "T_m" is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\% G + C)$, when a nucleic acid is in aqueous solution at 1 M NaCl (*See e.g.*, Anderson and Young, Quantitative Filter Hybridization, *in Nucleic Acid Hybridization* (1985). Other references include more sophisticated computations which take structural as well as sequence characteristics into account for the calculation of T_m.

As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds, under which nucleic acid hybridizations are conducted. With "high stringency" conditions, nucleic acid base pairing will occur only between nucleic acid fragments that have a high frequency of complementary base sequences. Thus, conditions of "weak" or "low" stringency are often required when it is desired that nucleic acids which are not completely complementary to one another be hybridized or annealed together.

The term "gene" refers to a DNA sequence that comprises control and coding sequences necessary for the production of a polypeptide or precursor. The polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired enzymatic activity is retained.

The term "wild-type" refers to a gene or gene product which has the characteristics of that gene or gene product when isolated from a naturally occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designed the "normal" or "wild-type" form of the gene. In contrast, the term "modified" or "mutant" refers to a gene or gene product which displays modifications in sequence and or functional

properties (*i.e.*, altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally-occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

The term "recombinant DNA vector" as used herein refers to DNA sequences containing a desired coding sequence and appropriate DNA sequences necessary for the expression of the operably linked coding sequence in a particular host organism. DNA sequences necessary for expression in procaryotes include a promoter, optionally an operator sequence, a ribosome binding site and possibly other sequences. Eukaryotic cells are known to utilize promoters, polyadenlyation signals and enhancers.

The term "LTR" as used herein refers to the long terminal repeat found at each end of a provirus (*i.e.*, the integrated form of a retrovirus). The LTR contains numerous regulatory signals including transcriptional control elements, polyadenylation signals and sequences needed for replication and integration of the viral genome. The viral LTR is divided into three regions called U3, R and U5.

The U3 region contains the enhancer and promoter elements. The U5 region contains the polyadenylation signals. The R (repeat) region separates the U3 and U5 regions and transcribed sequences of the R region appear at both the 5' and 3' ends of the viral RNA.

The term "oligonucleotide" as used herein is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably at least 5 nucleotides, more preferably at least about 10-15 nucleotides and more preferably at least about 15 to 30 nucleotides. The exact size will depend on many factors, which in turn depends on the ultimate function or use of the oligonucleotide. The oligonucleotide may be generated in any manner, including chemical synthesis, DNA replication, reverse transcription, or a combination thereof.

Because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage, an end of an oligonucleotide is referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends. A first region along a nucleic acid strand is said to be upstream of another region if the

3' end of the first region is before the 5' end of the second region when moving along a strand of nucleic acid in a 5' to 3' direction.

When two different, non-overlapping oligonucleotides anneal to different regions of the same linear complementary nucleic acid sequence, and the 3' end of one oligonucleotide points towards the 5' end of the other, the former may be called the "upstream" oligonucleotide and the latter the "downstream" oligonucleotide.

The term "primer" refers to an oligonucleotide which is capable of acting as a point of initiation of synthesis when placed under conditions in which primer extension is initiated. An oligonucleotide "primer" may occur naturally, as in a purified restriction digest or may be produced synthetically.

A primer is selected to be "substantially" complementary to a strand of specific sequence of the template. A primer must be sufficiently complementary to hybridize with a template strand for primer elongation to occur. A primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being substantially complementary to the strand. Non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the template to hybridize and thereby form a template primer complex for synthesis of the extension product of the primer.

"Hybridization" methods involve the annealing of a complementary sequence to the target nucleic acid (the sequence to be detected; the detection of this sequence may be by either direct or indirect means). The ability of two polymers of nucleic acid containing complementary sequences to find each other and anneal through base pairing interaction is a well-recognized phenomenon. The initial observations of the "hybridization" process by Marmur and Lane, *Proc. Natl. Acad. Sci. USA* 46:453 (1960) and Doty *et al.*, *Proc. Natl. Acad. Sci. USA* 46:461 (1960) have been followed by the refinement of this process into an essential tool of modern biology.

With regard to complementarity, it is important for some diagnostic applications to determine whether the hybridization represents complete or partial complementarity. For example, where it is desired to detect simply the presence or absence of pathogen DNA (such as from a virus, bacterium, fungi, mycoplasma, protozoan) it is only important that the hybridization method ensures hybridization when the relevant sequence is present; conditions can be selected where both partially complementary probes and completely complementary

probes will hybridize. Other diagnostic applications, however, may require that the hybridization method distinguish between partial and complete complementarity. It may be of interest to detect genetic polymorphisms. For example, human hemoglobin is composed, in part, of four polypeptide chains. Two of these chains are identical chains of 141 amino acids (alpha chains) and two of these chains are identical chains of 146 amino acids (beta chains). The gene encoding the beta chain is known to exhibit polymorphism. The normal allele encodes a beta chain having glutamic acid at the sixth position. The mutant allele encodes a beta chain having valine at the sixth position. This difference in amino acids has a profound (most profound when the individual is homozygous for the mutant allele) physiological impact known clinically as sickle cell anemia. It is well known that the genetic basis of the amino acid change involves a single base difference between the normal allele DNA sequence and the mutant allele DNA sequence.

The complement of a nucleic acid sequence as used herein refers to an oligonucleotide which, when aligned with the nucleic acid sequence such that the 5' end of one sequence is paired with the 3' end of the other, is in "antiparallel association." Certain bases not commonly found in natural nucleic acids may be included in the nucleic acids of the present invention and include, for example, inosine and 7-deazaguanine. Complementarity need not be perfect; stable duplexes may contain mismatched base pairs or unmatched bases. Those skilled in the art of nucleic acid technology can determine duplex stability empirically considering a number of variables including, for example, the length of the oligonucleotide, base composition and sequence of the oligonucleotide, ionic strength and incidence of mismatched base pairs.

Stability of a nucleic acid duplex is measured by the melting temperature, or "T_m." The T_m of a particular nucleic acid duplex under specified conditions is the temperature at which on average half of the base pairs have disassociated.

The term "label" as used herein refers to any atom or molecule which can be used to provide a detectable (preferably quantifiable) signal, and which can be attached to a nucleic acid or protein. Labels may provide signals detectable by fluorescence, radioactivity, colorimetry, gravimetry, X-ray diffraction or absorption, magnetism, enzymatic activity, and the like. A label may be a charged moeity (positive or negative charge) or alternatively, may be charge neutral.

The term "cleavage structure" as used herein, refers to a structure which is formed by the interaction of a probe oligonucleotide and a target nucleic acid to form a duplex, said

resulting structure being cleavable by a cleavage means, including but not limited to an enzyme. The cleavage structure is a substrate for specific cleavage by said cleavage means in contrast to a nucleic acid molecule which is a substrate for non-specific cleavage by agents such as phosphodiesterases which cleave nucleic acid molecules without regard to secondary structure (*i.e.*, no formation of a duplexed structure is required).

The term "cleavage means" as used herein refers to any means which is capable of cleaving a cleavage structure, including but not limited to enzymes. The cleavage means may include native DNAPs having 5' nuclease activity (*e.g.*, *Taq* DNA polymerase, *E. coli* DNA polymerase I) and, more specifically, modified DNAPs having 5' nuclease but lacking synthetic activity. The ability of 5' nucleases to cleave naturally occurring structures in nucleic acid templates (structure-specific cleavage) is useful to detect internal sequence differences in nucleic acids without prior knowledge of the specific sequence of the nucleic acid. In this manner, they are structure-specific enzymes. "Structure-specific nucleases" or "structure-specific enzymes" are enzymes which recognize specific secondary structures in a nucleic molecule and cleave these structures. The cleavage means of the invention cleave a nucleic acid molecule in response to the formation of cleavage structures; it is not necessary that the cleavage means cleave the cleavage structure at any particular location within the cleavage structure.

The cleavage means is not restricted to enzymes having solely 5' nuclease activity. The cleavage means may include nuclease activity provided from a variety of sources including the Cleavase® enzymes, the FEN-1 endonucleases (including RAD2 and XPG proteins), *Taq* DNA polymerase and *E. coli* DNA polymerase I.

The term "thermostable" when used in reference to an enzyme, such as a 5' nuclease, indicates that the enzyme is functional or active (*i.e.*, can perform catalysis) at an elevated temperature (*i.e.*, at about 55°C or higher).

The term "cleavage products" as used herein, refers to products generated by the reaction of a cleavage means with a cleavage structure (*i.e.*, the treatment of a cleavage structure with a cleavage means).

The term "target nucleic acid"refers to a nucleic acid molecule which contains a sequence which has at least partial complementarity with at least a probe oligonucleotide and may also have at least partial complementarity with an Invader™ oligonucleotide. The target nucleic acid may comprise single- or double-stranded DNA or RNA.

The term "probe oligonucleotide" refers to an oligonucleotide which interacts with a target nucleic acid to form a cleavage structure in the presence or absence of an Invader™ oligonucleotide. When annealed to the target nucleic acid, the probe oligonucleotide and target form a cleavage structure and cleavage occurs within the probe oligonucleotide. In the presence of an Invader™ oligonucleotide upstream of the probe oligonucleotide along the target nucleic acid will shift the site of cleavage within the probe oligonucleotide (relative to the site of cleavage in the absence of the Invader™).

The term "non-target cleavage product" refers to a product of a cleavage reaction which is not derived from the target nucleic acid. As discussed above, in the methods of the present invention, cleavage of the cleavage structure occurs within the probe oligonucleotide. The fragments of the probe oligonucleotide generated by this target nucleic acid-dependent cleavage are "non-target cleavage products."

The term "Invader™ oligonucleotide" refers to an oligonucleotide which contains sequences at its 3' end which are substantially the same as sequences located at the 5' end of a probe oligonucleotide; these regions will compete for hybridization to the same segment along a complementary target nucleic acid.

The term "substantially single-stranded" when used in reference to a nucleic acid substrate means that the substrate molecule exists primarily as a single strand of nucleic acid in contrast to a double-stranded substrate which exists as two strands of nucleic acid which are held together by inter-strand base pairing interactions.

The term "sequence variation" as used herein refers to differences in nucleic acid sequence between two nucleic acids. For example, a wild-type structural gene and a mutant form of this wild-type structural gene may vary in sequence by the presence of single base substitutions and/or deletions or insertions of one or more nucleotides. These two forms of the structural gene are said to vary in sequence from one another. A second mutant form of the structural gene may exist. This second mutant form is said to vary in sequence from both the wild-type gene and the first mutant form of the gene.

The term "liberating" as used herein refers to the release of a nucleic acid fragment from a larger nucleic acid fragment, such as an oligonucleotide, by the action of a 5' nuclease such that the released fragment is no longer covalently attached to the remainder of the oligonucleotide.

The term "K_m" as used herein refers to the Michaelis-Menten constant for an enzyme and is defined as the concentration of the specific substrate at which a given enzyme yields one-half its maximum velocity in an enzyme catalyzed reaction.

The term "nucleotide analog" as used herein refers to modified or non-naturally occurring nucleotides such as 7-deaza purines (*i.e.*, 7-deaza-dATP and 7-deaza-dGTP). Nucleotide analogs include base analogs and comprise modified forms of deoxyribonucleotides as well as ribonucleotides.

The term "polymorphic locus" is a locus present in a population which shows variation between members of the population (*i.e.*, the most common allele has a frequency of less than 0.95). In contrast, a "monomorphic locus" is a genetic locus at little or no variations seen between members of the population (generally taken to be a locus at which the most common allele exceeds a frequency of 0.95 in the gene pool of the population).

The term "microorganism" as used herein means an organism too small to be observed with the unaided eye and includes, but is not limited to bacteria, virus, protozoans, fungi, and ciliates.

The term "microbial gene sequences" refers to gene sequences derived from a microorganism.

The term "bacteria" refers to any bacterial species including eubacterial and archaebacterial species.

The term "virus" refers to obligate, ultramicroscopic, intracellular parasites incapable of autonomous replication (*i.e.*, replication requires the use of the host cell's machinery).

The term "multi-drug resistant" or multiple-drug resistant" refers to a microorganism which is resistant to more than one of the antibiotics or antimicrobial agents used in the treatment of said microorganism.

The term "sample" in the present specification and claims is used in its broadest sense. On the one hand it is meant to include a specimen or culture (*e.g.*, microbiological cultures). On the other hand, it is meant to include both biological and environmental samples.

Biological samples may be animal, including human, fluid, solid (*e.g.*, stool) or tissue, as well as liquid and solid food and feed products and ingredients such as dairy items, vegetables, meat and meat by-products, and waste. Biological samples may be obtained from all of the various families of domestic animals, as well as feral or wild animals, including, but not limited to, such animals as ungulates, bear, fish, lagamorphs, rodents, etc.

Environmental samples include environmental material such as surface matter, soil, water and industrial samples, as well as samples obtained from food and dairy processing instruments, apparatus, equipment, utensils, disposable and non-disposable items. These examples are not to be construed as limiting the sample types applicable to the present invention.

The term "source of target nucleic acid" refers to any sample which contains nucleic acids (RNA or DNA). Particularly preferred sources of target nucleic acids are biological samples including, but not limited to blood, saliva, cerebral spinal fluid, pleural fluid, milk, lymph, sputum and semen.

An oligonucleotide is said to be present in "excess" relative to another oligonucleotide (or target nucleic acid sequence) if that oligonucleotide is present at a higher molar concentration that the other oligonucleotide (or target nucleic acid sequence). When an oligonucleotide such as a probe oligonucleotide is present in a cleavage reaction in excess relative to the concentration of the complementary target nucleic acid sequence, the reaction may be used to indicate the amount of the target nucleic acid present. Typically, when present in excess, the probe oligonucleotide will be present at least a 100-fold molar excess; typically at least 1 pmole of each probe oligonucleotide would be used when the target nucleic acid sequence was present at about 10 fmoles or less.

A sample "suspected of containing" a first and a second target nucleic acid may contain either, both or neither target nucleic acid molecule.

The term "charge-balanced" oligonucleotide refers to an oligonucleotide (the input oligonucleotide in a reaction) which has been modified such that the modified oligonucleotide bears a charge, such that when the modified oligonucleotide is either cleaved (*i.e.*, shortened) or elongated, a resulting product bears a charge different from the input oligonucleotide (the "charge-unbalanced" oligonucleotide) thereby permitting separation of the input and reacted oligonucleotides on the basis of charge. The term "charge-balanced" does not imply that the modified or balanced oligonucleotide has a net neutral charge (although this can be the case). Charge-balancing refers to the design and modification of an oligonucleotide such that a specific reaction product generated from this input oligonucleotide can be separated on the basis of charge from the input oligonucleotide.

For example, in an Invader™-directed cleavage assay in which the probe oligonucleotide bears the sequence: 5'-TTCTTTTCACCAGCGAGACGGG-3' (*i.e.*, SEQ ID NO:50 without the modified bases) and cleavage of the probe occurs between the second and

third residues, one possible charge-balanced version of this oligonucleotide would be: 5'-Cy3-AminoT-Amino-TCTTTTCACCAGCGAGAC GGG-3'. This modified oligonucleotide bears a net negative charge. After cleavage, the following oligonucleotides are generated: 5'-Cy3-AminoT-Amino-T-3'and 5'-CTTTTCACCAGCGAGACGGG-3' (residues 3-22 of SEQ ID NO:50). 5'-Cy3-AminoT-Amino-T-3' bears a detectable moeity (the positively-charged Cy3 dye) and two amino-modified bases. The amino-modified bases and the Cy3 dye contribute positive charges in excess of the negative charges contributed by the phosphate groups and thus the 5'-Cy3-AminoT-Amino-T-3'oligonucleotide has a net positive charge. The other, longer cleavage fragment, like the input probe, bears a net negative charge. Because the 5'-Cy3-AminoT-Amino-T-3'fragment is separable on the basis of charge from the input probe (the charge-balanced oligonucleotide), it is referred to as a charge-unbalanced oligonucleotide. The longer cleavage product cannot be separated on the basis of charge from the input oligonucleotide as both oligonucleotides bear a net negative charge; thus, the longer cleavage product is not a charge-unbalanced oligonucleotide.

The term "net neutral charge" when used in reference to an oligonucleotide, including modified oligonucleotides, indicates that the sum of the charges present (*i.e*, R-NH^{3+} groups on thymidines, the N3 nitrogen of cytosine, presence or absence or phosphate groups, etc.) under the desired reaction conditions is essentially zero. An oligonucleotide having a net neutral charge would not migrate in an electrical field.

The term "net positive charge" when used in reference to an oligonucleotide, including modified oligonucleotides, indicates that the sum of the charges present (*i.e*, R-NH^{3+} groups on thymidines, the N3 nitrogen of cytosine, presence or absence or phosphate groups, etc.) under the desired reaction conditions is +1 or greater. An oligonucleotide having a net positive charge would migrate toward the negative electrode in an electrical field.

The term "net negative charge" when used in reference to an oligonucleotide, including modified oligonucleotides, indicates that the sum of the charges present (*i.e*, R-NH^{3+} groups on thymidines, the N3 nitrogen of cytosine, presence or absence or phosphate groups, etc.) under the desired reaction conditions is -1 or lower. An oligonucleotide having a net negative charge would migrate toward the positive electrode in an electrical field.

The term "polymerization means" refers to any agent capable of facilitating the addition of nucleoside triphosphates to an oligonucleotide. Preferred polymerization means comprise DNA polymerases.

The term "ligation means" refers to any agent capable of facilitating the ligation (*i.e.*, the formation of a phosphodiester bond between a 3'-OH and a 5'-P located at the termini of two strands of nucleic acid). Preferred ligation means comprise DNA ligases and RNA ligases.

The term "reactant" is used herein in its broadest sense. The reactant can comprise an enzymatic reactant, a chemical reactant or ultraviolet light (ultraviolet light, particularly short wavelength ultraviolet light is known to break oligonucleotide chains). Any agent capable of reacting with an oligonucleotide to either shorten (*i.e.*, cleave) or elongate the oligonucleotide is encompassed within the term "reactant."

The term "adduct" is used herein in its broadest sense to indicate any compound or element which can be added to an oligonucleotide. An adduct may be charged (positively or negatively) or may be charge neutral. An adduct may be added to the oligonucleotide via covalent or non-covalent linkages. Examples of adducts, include but are not limited to indodicarbocyanine dyes (*e.g.*, Cy3 and Cy5), amino-substituted nucleotides, ethidium bromide, ethidium homodimer, (1,3-propanediamino)propidium, (diethylenetriamino)propidium, thiazole orange, (N-N'-tetramethyl-1,3-propanediamino)propyl thiazole orange, (N-N'-tetramethyl-1,2-ethanediamino)propyl thiazole orange, thiazole orange-thiazole orange homodimer (TOTO), thiazole orande-thiazole blue heterodimer (TOTAB), thiazole orange-ethidium heterodimer 1 (TOED1), thiazole orange-ethidium heterodimer 2 (TOED2) and florescien-ethidium heterodimer (FED), psoralens, biotin, streptavidin, avidin, etc.

Where a first oligonucleotide is complementary to a region of a target nucleic acid and a second oligonucleotide has complementary to the same region (or a portion of this region) a "region of overlap" exists along the target nucleic acid. The degree of overlap will vary depending upon the nature of the complementarity (*See e.g.*, region "X" in Figs. 25 and 56 and the accompanying discussions).

As used herein, the term "purified" or "to purify" refers to the removal of contaminants from a sample. For example, recombinant Cleavase® nucleases are expressed in bacterial host cells and the nucleases are purified by the removal of host cell proteins; the percent of these recombinant nucleases is thereby increased in the sample.

The term "recombinant DNA molecule" as used herein refers to a DNA molecule which is comprised of segments of DNA joined together by means of molecular biological techniques.

The term "recombinant protein" or "recombinant polypeptide" as used herein refers to a protein molecule which is expressed from a recombinant DNA molecule.

As used herein the term "portion" when in reference to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from four amino acid residues to the entire amino acid sequence minus one amino acid.

"Nucleic acid sequence" as used herein refers to an oligonucleotide, nucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand. Similarly, "amino acid sequence" as used herein refers to peptide or protein sequence.

"Peptide nucleic acid" ("PNA") as used herein refers to a molecule which comprises an oligomer to which an amino acid residue, such as lysine, and an amino group have been added. These small molecules, also designated anti-gene agents, stop transcript elongation by binding to their complementary strand of nucleic acid (Nielsen *et al.*, Anticancer Drug Des. 8:53-63 [1993].)

As used herein, the terms "purified" or "substantially purified" refer to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated. An "isolated polynucleotide" or "isolated oligonucleotide" is therefore a substantially purified polynucleotide.

An isolated oligonucleotide (or polynucleotide) encoding a *Pyrococcus woesei* (*Pwo*) FEN-1 endonuclease having a region capable of hybridizing to SEQ ID NO:80 is an oligonucleotide containing sequences encoding at least the amino-terminal portion of *Pwo* FEN-1 endonuclease. An isolated oligonucleotide (or polynucleotide) encoding a *Pwo* FEN-1 endonuclease having a region capable of hybridizing to SEQ ID NO:81 is an oligonucleotide containing sequences encoding at least the carboxy-terminal portion of *Pwo* FEN-1 endonuclease. An isolated oligonucleotide (or polynucleotide) encoding a *Pwo* FEN-1 endonuclease having a region capable of hybridizing to SEQ ID NOS:82 and 83 is an oligonucleotide containing sequences encoding at least portions of *Pwo* FEN-1 endonuclease protein located internal to either the amino or carboxy-termini of the *Pwo* FEN-1 endonuclease protein.

As used herein, the term "fusion protein" refers to a chimeric protein containing the protein of interest (*i.e.*, Cleavase® BN/thrombin nuclease and portions or fragments thereof)

joined to an exogenous protein fragment (the fusion partner which consists of a non-Cleavase® BN/thrombin nuclease protein). The fusion partner may enhance solubility of recombinant chimeric protein (*e.g.*, the Cleavase® BN/thrombin nuclease) as expressed in a host cell, may provide an affinity tag (*e.g.*, a his-tag) to allow purification of the recombinant fusion protein from the host cell or culture supernatant, or both. If desired, the fusion protein may be removed from the protein of interest (*e.g.*, Cleavase® BN/thrombin nuclease or fragments thereof) by a variety of enzymatic or chemical means known to the art.

The term "purified *Pfu* FEN-1 endonuclease having a molecular weight of about 38.7 kilodaltons" refers to a FEN-1 endonuclease isolated from *Pyrococcus woesei* which has a molecular weight on SDS-PAGE gels of about 38.7 kDa when the SDS-PAGE is conducted under the conditions described in Ex. 28. Those skilled in the art understand that the same protein preparation applied to separate gels of apparently the same composition can yield estimated molecular weights which vary somewhat from one another (approximately 5-15%).

The term "continuous strand of nucleic acid" as used herein is means a strand of nucleic acid that has a continuous, covalently linked, backbone structure, without nicks or other disruptions. The disposition of the base portion of each nucleotide, whether base-paired, single-stranded or mismatched, is not an element in the definition of a continuous strand. The backbone of the continuous strand is not limited to the ribose-phosphate or deoxyribose-phosphate compositions that are found in naturally occurring, unmodified nucleic acids. A nucleic acid of the present invention may comprise modifications in the structure of the backbone, including but not limited to phosphorothioate residues, phosphonate residues, 2' substituted ribose residues (*e.g.*, 2'-O-methyl ribose) and alternative sugar (*e.g.*, arabinose) containing residues.

The term "continuous duplex" as used herein refers to a region of double stranded nucleic acid in which there is no disruption in the progression of basepairs within the duplex (*i.e.*, the base pairs along the duplex are not distorted to accommodate a gap, bulge or mismatch with the confines of the region of continuous duplex). As used herein the term refers only to the arrangement of the basepairs within the duplex, without implication of continuity in the backbone portion of the nucleic acid strand. Duplex nucleic acids with uninterrupted basepairing, but with nicks in one or both strands are within the definition of a continuous duplex.

The term "duplex" refers to the state of nucleic acids in which the base portions of the nucleotides on one strand are bound through hydrogen bonding the their complementary bases

arrayed on a second strand. The condition of being in a duplex form reflects on the state of the bases of a nucleic acid. By virtue of base pairing, the strands of nucleic acid also generally assume the tertiary structure of a double helix, having a major and a minor groove. The assumption of the helical form is implicit in the act of becoming duplexed.

The term "duplex dependent protein binding" refers to the binding of proteins to nucleic acid that is dependent on the nucleic acid being in a duplex,or helical form.

The term "duplex dependent protein binding sites or regions" as used herein refers to discrete regions or sequences within a nucleic acid that are bound with particular affinity by specific duplex-dependent nucleic acid binding proteins. This is in contrast to the generalized duplex-dependent binding of proteins that are not site-specific, such as the histone proteins that bind chromatin with little reference to specific sequences or sites.

The term "protein binding region" as used herein refers to a nucleic acid region identified by a sequence or structure as binding to a particular protein or class of proteins. It is within the scope of this definition to include those regions that contain sufficient genetic information to allow identifications of the region by comparison to known sequences, but which might not have the requisite structure for actual binding (*e.g.*, a single strand of a duplex-depending nucleic acid binding protein site). As used herein "protein binding region" excludes restriction endonuclease binding regions.

The term "complete double stranded protein binding region" as used herein refers to the minimum region of continuous duplex required to allow binding or other activity of a duplex-dependent protein. This definition is intended to encompass the observation that some duplex dependent nucleic acid binding proteins can interact with full activity with regions of duplex that may be shorter than a canonical protein binding region as observed in one or the other of the two single strands. In other words, one or more nucleotides in the region may be allowed to remain unpaired without suppressing binding. As used here in, the term "complete double stranded binding region" refers to the minimum sequence that will accommodate the binding function. Because some such regions can tolerate non-duplex sequences in multiple places, although not necessarily simultaneously, a single protein binding region might have several shorter sub-regions that, when duplexed, will be fully competent for protein binding.

The term "template" refers to a strand of nucleic acid on which a complementary copy is built from nucleoside triphosphates through the activity of a template-dependent nucleic acid polymerase. Within a duplex the template strand is, by convention, depicted

and described as the "bottom" strand. Similarly, the non-template strand is often depicted and described as the "top" strand.

The term "template-dependent RNA polymerase" refers to a nucleic acid polymerase that creates new RNA strands through the copying of a template strand as described above and which does not synthesize RNA in the absence of a template. This is in contrast to the activity of the template-independent nucleic acid polymerases that synthesize or extend nucleic acids without reference to a template, such as terminal deoxynucleotidyl transferase, or Poly A polymerase.

The term "Arrestor™" refers to an agent added to or included in an invasive cleavage reaction in order to stop one or more reaction components from participating in a subsequent action or reaction. This may be done by sequestering or inactivating some reaction component (*e.g.*, by binding or base-pairing a nucleic acid component, or by binding to a protein component). The term "Arrestor™ oligonucleotide" refers to an oligonucleotide included in an invasive cleavage reaction in order to stop or arrest one or more aspects of any reaction (*i.e.*, the first reaction and/or any subsequent reactions or actions; it is not intended that the Arrestor™ oligonucleotide be limited to any particular reaction or reaction step). This may be done by sequestering some reaction component (*e.g.*, base-pairing to another nucleic acid, or binding to a protein component). However, it is not intended that the term be so limited as to just situations in which a reaction component is sequestered.

DESCRIPTION OF THE INVENTION

The present invention relates to methods and compositions for treating nucleic acid, and in particular, methods and compositions for detection and characterization of nucleic acid sequences and sequence changes.

The present invention relates to means for cleaving a nucleic acid cleavage structure in a site-specific manner. In particular, the present invention relates to a cleaving enzyme having 5' nuclease activity without interfering nucleic acid synthetic ability.

This invention provides 5' nucleases derived from thermostable DNA polymerases which exhibit altered DNA synthetic activity from that of native thermostable DNA polymerases. The 5' nuclease activity of the polymerase is retained while the synthetic activity is reduced or absent. Such 5' nucleases are capable of catalyzing the structure-specific cleavage of nucleic acids in the absence of interfering synthetic activity. The lack of

synthetic activity during a cleavage reaction results in nucleic acid cleavage products of uniform size.

The novel properties of the nucleases of the invention form the basis of a method of detecting specific nucleic acid sequences. This method relies upon the amplification of the detection molecule rather than upon the amplification of the target sequence itself as do existing methods of detecting specific target sequences.

DNA polymerases (DNAPs), such as those isolated from *E. coli* or from thermophilic bacteria of the genus *Thermus*, are enzymes that synthesize new DNA strands. Several of the known DNAPs contain associated nuclease activities in addition to the synthetic activity of the enzyme.

Some DNAPs are known to remove nucleotides from the 5' and 3' ends of DNA chains (Kornberg, *DNA Replication,* W.H. Freeman and Co., San Francisco, pp. 127-139 [1980]). These nuclease activities are usually referred to as 5' exonuclease and 3' exonuclease activities, respectively. For example, the 5' exonuclease activity located in the N-terminal domain of several DNAPs participates in the removal of RNA primers during lagging strand synthesis during DNA replication and the removal of damaged nucleotides during repair. Some DNAPs, such as the *E. coli* DNA polymerase (DNAPEc1), also have a 3' exonuclease activity responsible for proof-reading during DNA synthesis (Kornberg, *supra*).

A DNAP isolated from *Thermus aquaticus*, termed *Taq* DNA polymerase (DNAP*Taq*), has a 5' exonuclease activity, but lacks a functional 3' exonucleolytic domain (Tindall and Kunkell, *Biochem.*, 27:6008 [1988]). Derivatives of DNAPEc1 and DNAP*Taq*, respectively called the Klenow and Stoffel fragments, lack 5' exonuclease domains as a result of enzymatic or genetic manipulations (Brutlag *et al.*, *Biochem. Biophys. Res. Commun.* 37:982 [1969]; Erlich *et al.*, *Science* 252:1643 [1991]; Setlow and Kornberg, *J. Biol. Chem.* 247:232 [1972]).

The 5' exonuclease activity of DNAP*Taq* was reported to require concurrent synthesis (Gelfand, *in PCR Technology - Principles and Applications for DNA Amplification* Erlich, (Ed.), Stockton Press, New York, p. 19 [1989]). Although mononucleotides predominate among the digestion products of the 5' exonucleases of DNAP*Taq* and DNAPEc1, short oligonucleotides (≤ 12 nucleotides) can also be observed implying that these so-called 5' exonucleases can function endonucleolytically (Setlow, *supra*; Holland *et al.*, *Proc. Natl. Acad. Sci.* USA 88:7276 [1991]).

In WO 92/06200, Gelfand *et al.* show that the preferred substrate of the 5' exonuclease activity of the thermostable DNA polymerases is displaced single-stranded DNA. Hydrolysis of the phosphodiester bond occurs between the displaced single-stranded DNA and the double-helical DNA with the preferred exonuclease cleavage site being a phosphodiester bond in the double helical region. Thus, the 5' exonuclease activity usually associated with DNAPs is a structure-dependent single-stranded endonuclease and is more properly referred to as a 5' nuclease. Exonucleases are enzymes which cleave nucleotide molecules from the ends of the nucleic acid molecule. Endonucleases, on the other hand, are enzymes which cleave the nucleic acid molecule at internal rather than terminal sites. The nuclease activity associated with some thermostable DNA polymerases cleaves endonucleolytically but this cleavage requires contact with the 5' end of the molecule being cleaved. Therefore, these nucleases are referred to as 5' nucleases.

When a 5' nuclease activity is associated with a eubacterial Type A DNA polymerase, it is found in the one-third N-terminal region of the protein as an independent functional domain. The C-terminal two-thirds of the molecule constitute the polymerization domain which is responsible for the synthesis of DNA. Some Type A DNA polymerases also have a 3' exonuclease activity associated with the two-third C-terminal region of the molecule.

The 5' exonuclease activity and the polymerization activity of DNAPs have been separated by proteolytic cleavage or genetic manipulation of the polymerase molecule. To date thermostable DNAPs have been modified to remove or reduce the amount of 5' nuclease activity while leaving the polymerase activity intact.

The Klenow or large proteolytic cleavage fragment of DNAPEc1 contains the polymerase and 3' exonuclease activity but lacks the 5' nuclease activity. The Stoffel fragment of DNAP*Taq* (DNAPStf) lacks the 5' nuclease activity due to a genetic manipulation which deleted the N-terminal 289 amino acids of the polymerase molecule (Erlich *et al.*, *Science* 252:1643 [1991]). WO 92/06200 describes a thermostable DNAP with an altered level of 5' to 3' exonuclease. U.S. Patent No. 5,108,892 describes a *Thermus aquaticus* DNAP without a 5' to 3' exonuclease. However, the art of molecular biology lacks a thermostable DNA polymerase with a lessened amount of synthetic activity.

The present invention provides 5' nucleases derived from thermostable Type A DNA polymerases that retain 5' nuclease activity but have reduced or absent synthetic activity. The ability to uncouple the synthetic activity of the enzyme from the 5' nuclease activity proves

that the 5' nuclease activity does not require concurrent DNA synthesis as was previously reported (Gelfand, *PCR Technology, supra*).

The description of the invention is divided into: I. Generation of 5' Nucleases Derived From Thermostable DNA Polymerases; II. Detection of Specific Nucleic Acid Sequences Using 5' Nucleases in an Invader™-Directed Cleavage Assay; III. A Comparison Of Invasive Cleavage And Primer-Directed Cleavage; IV. Fractionation Of Specific Nucleic Acids By Selective Charge Reversal; V. Invader™-Directed Cleavage Using Miniprobes And Mid-Range Probes; VI. Signal Enhancement By Tailing Of Reaction Products In The Invader™-Directed Cleavage Assay; VII. Improved Enzymes For Use In Invader™-Directed Cleavage Reactions; VIII. Signal Enhancement By Completion Of An Activated Protein Binding Site; IX. Signal Enhancement By Incorporating The Products Of An Invasive Cleavage Reaction Into A Subsequent Invasive Cleavage Reaction; X. Detection of Human Cytomegalovirus Viral DNA By Invasive Cleavage; and XI. Effect of Arrestor™ Oligonucleotides on Signal and Background in Sequential Invasive Cleavage Reactions.

I. Generation Of 5' Nucleases From Thermostable DNA Polymerases

The genes encoding Type A DNA polymerases share about 85% homology to each other on the DNA sequence level. Preferred examples of thermostable polymerases include those isolated from *Thermus aquaticus*, *Thermus flavus*, and *Thermus thermophilus*. However, other thermostable Type A polymerases which have 5' nuclease activity are also suitable. Figs. 1 and 2 compare the nucleotide and amino acid sequences of the three above mentioned polymerases. In Figs. 1 and 2, the consensus or majority sequence derived from a comparison of the nucleotide (Fig. 1) or amino acid (Fig. 2) sequence of the three thermostable DNA polymerases is shown on the top line. A dot appears in the sequences of each of these three polymerases whenever an amino acid residue in a given sequence is identical to that contained in the consensus amino acid sequence. Dashes are used to introduce gaps in order to maximize alignment between the displayed sequences. When no consensus nucleotide or amino acid is present at a given position, an "X" is placed in the consensus sequence. SEQ ID NOS:1-3 display the nucleotide sequences and SEQ ID NOS:4-6 display the amino acid sequences of the three wild-type polymerases. SEQ ID NO:1 corresponds to the nucleic acid sequence of the wild type *Thermus aquaticus* DNA polymerase gene isolated from the YT-1 strain (Lawyer *et al.*, *J. Biol. Chem.*, 264:6427 [1989]). SEQ ID NO:2 corresponds to the nucleic acid sequence of the wild type *Thermus*

flavus DNA polymerase gene (Akhmetzjanov and Vakhitov, *Nucl. Acids Res.*, 20:5839 [1992]). SEQ ID NO:3 corresponds to the nucleic acid sequence of the wild type *Thermus thermophilus* DNA polymerase gene (Gelfand *et al.*, WO 91/09950 [1991]). SEQ ID NOS:7-8 depict the consensus nucleotide and amino acid sequences, respectively for the above three DNAPs (also shown on the top row in Figs. 2 and 3).

The 5' nucleases of the invention derived from thermostable polymerases have reduced synthetic ability, but retain substantially the same 5' exonuclease activity as the native DNA polymerase. The term "substantially the same 5' nuclease activity" as used herein means that the 5' nuclease activity of the modified enzyme retains the ability to function as a structure-dependent single-stranded endonuclease but not necessarily at the same rate of cleavage as compared to the unmodified enzyme. Type A DNA polymerases may also be modified so as to produce an enzyme which has increases 5' nuclease activity while having a reduced level of synthetic activity. Modified enzymes having reduced synthetic activity and increased 5' nuclease activity are also envisioned by the present invention.

By the term "reduced synthetic activity" as used herein it is meant that the modified enzyme has less than the level of synthetic activity found in the unmodified or "native" enzyme. The modified enzyme may have no synthetic activity remaining or may have that level of synthetic activity that will not interfere with the use of the modified enzyme in the detection assay described below. The 5' nucleases of the present invention are advantageous in situations where the cleavage activity of the polymerase is desired, but the synthetic ability is not (such as in the detection assay of the invention).

As noted above, it is not intended that the invention be limited by the nature of the alteration necessary to render the polymerase synthesis deficient. The present invention contemplates a variety of methods, including but not limited to: 1) proteolysis; 2) recombinant constructs (including mutants); and 3) physical and/or chemical modification and/or inhibition.

1. Proteolysis

Thermostable DNA polymerases having a reduced level of synthetic activity are produced by physically cleaving the unmodified enzyme with proteolytic enzymes to produce fragments of the enzyme that are deficient in synthetic activity but retain 5' nuclease activity. Following proteolytic digestion, the resulting fragments are separated by standard chromatographic techniques and assayed for the ability to synthesize DNA and to act as a 5'

nuclease. The assays to determine synthetic activity and 5' nuclease activity are described below.

2. Recombinant Constructs

The examples below describe a preferred method for creating a construct encoding a 5' nuclease derived from a thermostable DNA polymerase. As the Type A DNA polymerases are similar in DNA sequence, the cloning strategies employed for the *Thermus aquaticus* and *flavus* polymerases are applicable to other thermostable Type A polymerases. In general, a thermostable DNA polymerase is cloned by isolating genomic DNA using molecular biological methods from a bacteria containing a thermostable Type A DNA polymerase. This genomic DNA is exposed to primers which are capable of amplifying the polymerase gene by PCR.

This amplified polymerase sequence is then subjected to standard deletion processes to delete the polymerase portion of the gene. Suitable deletion processes are described below in the examples.

The example below discusses the strategy used to determine which portions of the DNAP*Taq* polymerase domain could be removed without eliminating the 5' nuclease activity. Deletion of amino acids from the protein can be done either by deletion of the encoding genetic material, or by introduction of a translational stop codon by mutation or frame shift. In addition, proteolytic treatment of the protein molecule can be performed to remove segments of the protein.

In the examples below, specific alterations of the *Taq* gene were: a deletion between nucleotides 1601 and 2502 (the end of the coding region), a 4 nucleotide insertion at position 2043, and deletions between nucleotides 1614 and 1848 and between nucleotides 875 and 1778 (numbering is as in SEQ ID NO:1). These modified sequences are described below in the examples and at SEQ ID NOS:9-12.

Those skilled in the art understand that single base pair changes can be innocuous in terms of enzyme structure and function. Similarly, small additions and deletions can be present without substantially changing the exonuclease or polymerase function of these enzymes.

Other deletions are also suitable to create the 5' nucleases of the present invention. It is preferable that the deletion decrease the polymerase activity of the 5' nucleases to a level at which synthetic activity will not interfere with the use of the 5' nuclease in the detection

assay of the invention. Most preferably, the synthetic ability is absent. Modified polymerases are tested for the presence of synthetic and 5' nuclease activity as in assays described below. Thoughtful consideration of these assays allows for the screening of candidate enzymes whose structure is heretofore as yet unknown. In other words, construct "X" can be evaluated according to the protocol described below to determine whether it is a member of the genus of 5' nucleases of the present invention as defined functionally, rather than structurally.

In the example below, the PCR product of the amplified *Thermus aquaticus* genomic DNA did not have the identical nucleotide structure of the native genomic DNA and did not have the same synthetic ability of the original clone. Base pair changes which result due to the infidelity of DNAP*Taq* during PCR amplification of a polymerase gene are also a method by which the synthetic ability of a polymerase gene may be inactivated. The examples below and Figs. 3A and 4A indicate regions in the native *Thermus aquaticus* and *flavus* DNA polymerases likely to be important for synthetic ability. There are other base pair changes and substitutions that will likely also inactivate the polymerase.

It is not necessary, however, that one start out the process of producing a 5' nuclease from a DNA polymerase with such a mutated amplified product. This is the method by which the examples below were performed to generate the synthesis-deficient DNAP*Taq* mutants, but it is understood by those skilled in the art that a wild-type DNA polymerase sequence may be used as the starting material for the introduction of deletions, insertion and substitutions to produce a 5' nuclease. For example, to generate the synthesis-deficient DNAPTfl mutant, the primers listed in SEQ ID NOS:13-14 were used to amplify the wild type DNA polymerase gene from *Thermus flavus* strain AT-62. The amplified polymerase gene was then subjected to restriction enzyme digestion to delete a large portion of the domain encoding the synthetic activity.

The present invention contemplates that the nucleic acid construct of the present invention be capable of expression in a suitable host. Those in the art know methods for attaching various promoters and 3' sequences to a gene structure to achieve efficient expression. The examples below disclose two suitable vectors and six suitable vector constructs. Of course, there are other promoter/vector combinations that would be suitable. It is not necessary that a host organism be used for the expression of the nucleic acid constructs of the invention. For example, expression of the protein encoded by a nucleic acid construct may be achieved through the use of a cell-free in vitro transcription/translation

system. An example of such a cell-free system is the commercially available TnT™ Coupled Reticulocyte Lysate System (Promega).

Once a suitable nucleic acid construct has been made, the 5' nuclease may be produced from the construct. The examples below and standard molecular biological teachings enable one to manipulate the construct by different suitable methods.

Once the 5' nuclease has been expressed, the polymerase is tested for both synthetic and nuclease activity as described below.

3. Physical And/Or Chemical Modification And/Or Inhibition

The synthetic activity of a thermostable DNA polymerase may be reduced by chemical and/or physical means. In one embodiment, the cleavage reaction catalyzed by the 5' nuclease activity of the polymerase is run under conditions which preferentially inhibit the synthetic activity of the polymerase. The level of synthetic activity need only be reduced to that level of activity which does not interfere with cleavage reactions requiring no significant synthetic activity.

As shown in the examples below, concentrations of Mg^{++} greater than 5 mM inhibit the polymerization activity of the native DNAP*Taq*. The ability of the 5' nuclease to function under conditions where synthetic activity is inhibited is tested by running the assays for synthetic and 5' nuclease activity, described below, in the presence of a range of Mg^{++} concentrations (5 to 10 mM). The effect of a given concentration of Mg^{++} is determined by quantitation of the amount of synthesis and cleavage in the test reaction as compared to the standard reaction for each assay.

The inhibitory effect of other ions, polyamines, denaturants, such as urea, formamide, dimethylsulfoxide, glycerol and non-ionic detergents (Triton X-100 and Tween-20), nucleic acid binding chemicals such as, actinomycin D, ethidium bromide and psoralens, are tested by their addition to the standard reaction buffers for the synthesis and 5' nuclease assays. Those compounds having a preferential inhibitory effect on the synthetic activity of a thermostable polymerase are then used to create reaction conditions under which 5' nuclease activity (cleavage) is retained while synthetic activity is reduced or eliminated.

Physical means may be used to preferentially inhibit the synthetic activity of a polymerase. For example, the synthetic activity of thermostable polymerases is destroyed by exposure of the polymerase to extreme heat (typically 96 to 100°C) for extended periods of

time (greater than or equal to 20 minutes). While these are minor differences with respect to the specific heat tolerance for each of the enzymes, these are readily determined. Polymerases are treated with heat for various periods of time and the effect of the heat treatment upon the synthetic and 5' nuclease activities is determined.

II. Detection Of Specific Nucleic Acid Sequences Using 5' Nucleases In An Invader™-Directed Cleavage Assay

The present invention provides means for forming a nucleic acid cleavage structure which is dependent upon the presence of a target nucleic acid and cleaving the nucleic acid cleavage structure so as to release distinctive cleavage products. 5' nuclease activity is used to cleave the target-dependent cleavage structure and the resulting cleavage products are indicative of the presence of specific target nucleic acid sequences in the sample.

The present invention further provides assays in which the target nucleic acid is reused or recycled during multiple rounds of hybridization with oligonucleotide probes and cleavage without the need to use temperature cycling (*i.e.*, for periodic denaturation of target nucleic acid strands) or nucleic acid synthesis (*i.e.*, for the displacement of target nucleic acid strands). Through the interaction of the cleavage means (*e.g.*, a 5' nuclease) an upstream oligonucleotide, the cleavage means can be made to cleave a downstream oligonucleotide at an internal site in such a way that the resulting fragments of the downstream oligonucleotide dissociate from the target nucleic acid, thereby making that region of the target nucleic acid available for hybridization to another, uncleaved copy of the downstream oligonucleotide.

As illustrated in Fig. 25, the methods of the present invention employ at least a pair of oligonucleotides that interact with a target nucleic acid to form a cleavage structure for a structure-specific nuclease. More specifically, the cleavage structure comprises: i) a target nucleic acid that may be either single-stranded or double-stranded (when a double-stranded target nucleic acid is employed, it may be rendered single stranded (*e.g.*, by heating); ii) a first oligonucleotide, termed the "probe," which defines a first region of the target nucleic acid sequence by being the complement of that region (regions X and Z of the target as shown in Fig. 25); and iii) a second oligonucleotide, termed the "Invader™," the 5' part of which defines a second region of the same target nucleic acid sequence (regions Y and X in Fig. 25), adjacent to and downstream of the first target region (regions X and Z), and the second part of which overlaps into the region defined by the first oligonucleotide (region X depicts the region of overlap). The resulting structure is diagrammed in Fig. 25.

While not limiting the invention or the instant discussion to any particular mechanism of action, the diagram in Fig. 25 represents the effect on the site of cleavage caused by this type of arrangement of a pair of oligonucleotides. The design of such a pair of oligonucleotides is described below in detail. In Fig. 25, the 3' ends of the nucleic acids (*i.e.*, the target and the oligonucleotides) are indicated by the use of the arrowheads on the ends of the lines depicting the strands of the nucleic acids (and where space permits, these ends are also labelled "3'"). It is readily appreciated that the two oligonucleotides (the Invader™ and the probe) are arranged in a parallel orientation relative to one another, while the target nucleic acid strand is arranged in an anti-parallel orientation relative to the two oligonucleotides. Further it is clear that the Invader™ oligonucleotide is located upstream of the probe oligonucleotide and that with respect to the target nucleic acid strand, region Z is upstream of region X and region X is upstream of region Y (that is region Y is downstream of region X and region X is downstream of region Z). Regions of complementarity between the opposing strands are indicated by the short vertical lines. While not intended to indicate the precise location of the site(s) of cleavage, the area to which the site of cleavage within the probe oligonucleotide is shifted by the presence of the Invader™ oligonucleotide is indicated by the solid vertical arrowhead. An alternative representation of the target/Invader™/probe cleavage structure is shown in Fig. 28C. Neither diagram (*i.e.*, Fig. 25 or Fig. 28C) is intended to represent the actual mechanism of action or physical arrangement of the cleavage structure and further it is not intended that the method of the present invention be limited to any particular mechanism of action.

It can be considered that the binding of these oligonucleotides divides the target nucleic acid into three distinct regions: one region that has complementarity to only the probe (shown as "Z"); one region that has complementarity only to the Invader™ (shown as "Y"); and one region that has complementarity to both oligonucleotides (shown as "X").

Design of these oligonucleotides (*i.e.*, the Invader™ and the probe) is accomplished using practices which are standard in the art. For example, sequences that have self complementarity, such that the resulting oligonucleotides would either fold upon themselves, or hybridize to each other at the expense of binding to the target nucleic acid, are generally avoided.

One consideration in choosing a length for these oligonucleotides is the complexity of the sample containing the target nucleic acid. For example, the human genome is approximately 3×10^9 basepairs in length. Any 10 nucleotide sequence will appear with a

frequency of $1:4^{10}$, or 1:1048,576 in a random string of nucleotides, which would be approximately 2,861 times in 3 billion basepairs. Clearly an oligonucleotide of this length would have a poor chance of binding uniquely to a 10 nucleotide region within a target having a sequence the size of the human genome. If the target sequence were within a 3 kb plasmid, however, such an oligonucleotide might have a very reasonable chance of binding uniquely. By this same calculation it can be seen that an oligonucleotide of 16 nucleotides (*i.e.*, a 16-mer) is the minimum length of a sequence which is mathematically likely to appear once in 3×10^9 basepairs.

A second consideration in choosing oligonucleotide length is the temperature range in which the oligonucleotides will be expected to function. A 16-mer of average base content (50% G-C basepairs) will have a calculated T_m (the temperature at which 50% of the sequence is dissociated) of about 41°C, depending on, among other things, the concentration of the oligonucleotide and its target, the salt content of the reaction and the precise order of the nucleotides. As a practical matter, longer oligonucleotides are usually chosen to enhance the specificity of hybridization. Oligonucleotides 20 to 25 nucleotides in length are often used as they are highly likely to be specific if used in reactions conducted at temperatures which are near their T_ms (within about 5° of the T_m). In addition, with calculated T_ms in the range of 50° to 70°C, such oligonucleotides (*i.e*, 20 to 25-mers) are appropriately used in reactions catalyzed by thermostable enzymes, which often display optimal activity near this temperature range.

The maximum length of the oligonucleotide chosen is also based on the desired specificity. One must avoid choosing sequences that are so long that they are either at a high risk of binding stably to partial complements, or that they cannot easily be dislodged when desired (*e.g.*, failure to disassociate from the target once cleavage has occurred).

The first step of design and selection of the oligonucleotides for the Invader™-directed cleavage is in accordance with these sample general principles. Considered as sequence-specific probes individually, each oligonucleotide may be selected according to the guidelines listed above. That is to say, each oligonucleotide will generally be long enough to be reasonably expected to hybridize only to the intended target sequence within a complex sample, usually in the 20 to 40 nucleotide range. Alternatively, because the Invader™-directed cleavage assay depends upon the concerted action of these oligonucleotides, the composite length of the 2 oligonucleotides which span/bind to the X, Y, Z regions may be selected to fall within this range, with each of the individual oligonucleotides being in

approximately the 13 to 17 nucleotide range. Such a design might be employed if a non-thermostable cleavage means were employed in the reaction, requiring the reactions to be conducted at a lower temperature than that used when thermostable cleavage means are employed. In some instances, it may be desirable to have these oligonucleotides bind multiple times within a target nucleic acid (*e.g.*, which bind to multiple variants or multiple similar sequences within a target). It is not intended that the method of the present invention be limited to any particular size of the probe or Invader™ oligonucleotide.

The second step of designing an oligonucleotide pair for this assay is to choose the degree to which the upstream "Invader™" oligonucleotide sequence will overlap into the downstream "probe" oligonucleotide sequence, and consequently, the sizes into which the probe will be cleaved. A key feature of this assay is that the probe oligonucleotide can be made to "turn over," that is to say cleaved probe can be made to depart to allow the binding and cleavage of other copies of the probe molecule, without the requirements of thermal denaturation or displacement by polymerization. While in one embodiment of this assay probe turnover may be facilitated by an exonucleolytic digestion by the cleavage agent, it is central to the present invention that the turnover does not <u>require</u> this exonucleolytic activity.

Choosing The Amount Of Overlap (Length Of The X Region)

One way of accomplishing such turnover can be envisioned by considering the diagram in Fig. 25. It can be seen that the T_m of each oligonucleotide will be a function of the full length of that oligonucleotide (*i.e.*, the T_m of the Invader™ = $T_{m(Y+X)}$, and the T_m of the probe = $T_{m(X+Y)}$ for the probe). When the probe is cleaved the X region is released, leaving the Z section. If the T_m of Z is less than the reaction temperature, and the reaction temperature is less than the $T_{m(X+Z)}$, then cleavage of the probe will lead to the departure of Z, thus allowing a new (X+Z) to hybridize. It can be seen from this example that the X region must be sufficiently long that the release of X will drop the T_m of the remaining probe section below the reaction temperature: a G-C rich X section may be much shorter than an A-T rich X section and still accomplish this stability shift.

Designing Oligonucleotides Which Interact With The Y And Z Regions

If the binding of the Invader™ oligonucleotide to the target is more stable than the binding of the probe (*e.g.*, if it is long, or is rich in G-C basepairs in the Y region), then the copy of X associated with the Invader™ may be favored in the competition for binding to the

X region of the target, and the probe may consequently hybridize inefficiently, and the assay may give low signal. Alternatively, if the probe binding is particularly strong in the Z region, the Invader™ will still cause internal cleavage, because this is mediated by the enzyme, but portion of the probe oligonucleotide bound to the Z region may not dissociate at the reaction temperature, turnover may be poor, and the assay may again give low signal.

It is clearly beneficial for the portions of the oligonucleotide which interact with the Y and Z regions so be similar in stability (*i.e.*, they must have similar melting temperatures). This is not to say that these regions must be the same length. As noted above, in addition to length, the melting temperature will also be affected by the base content and the specific sequence of those bases. The specific stability designed into the Invader™ and probe sequences will depend on the temperature at which one desires to perform the reaction.

This discussion is intended to illustrate that (within the basic guidelines for oligonucleotide specificity discussed above) it is the <u>balance</u> achieved between the stabilities of the probe and Invader™ sequences and their X and Y component sequences, rather than the absolute <u>values</u> of these stabilities, that is the chief consideration in the selection of the probe and Invader™ sequences.

Design Of The Reaction Conditions

Target nucleic acids that may be analyzed using the methods of the present invention which employ a 5' nuclease as the cleavage means include many types of both RNA and DNA. Such nucleic acids may be obtained using standard molecular biological techniques. For example, nucleic acids (RNA or DNA) may be isolated from a tissue sample (*e.g*, a biopsy specimen), tissue culture cells, samples containing bacteria and/or viruses (including cultures of bacteria and/or viruses), etc. The target nucleic acid may also be transcribed *in vitro* from a DNA template or may be chemically synthesized or generated in a PCR. Furthermore, nucleic acids may be isolated from an organism, either as genomic material or as a plasmid or similar extrachromosomal DNA, or they may be a fragment of such material generated by treatment with a restriction endonuclease or other cleavage agents or it may be synthetic.

Assembly of the target, probe, and Invader™ nucleic acids into the cleavage reaction of the present invention uses principles commonly used in the design of oligonucleotide base enzymatic assays, such as dideoxynucleotide sequencing and polymerase chain reaction (PCR). As is done in these assays, the oligonucleotides are provided in sufficient excess that

the rate of hybridization to the target nucleic acid is very rapid. These assays are commonly performed with 50 fmoles to 2 pmoles of each oligonucleotide per µl of reaction mixture. In the Examples described herein, amounts of oligonucleotides ranging from 250 fmoles to 5 pmoles per µl of reaction volume were used. These values were chosen for the purpose of ease in demonstration and are not intended to limit the performance of the present invention to these concentrations. Other (*e.g.*, lower) oligonucleotide concentrations commonly used in other molecular biological reactions are also contemplated.

It is desirable that an Invader™ oligonucleotide be immediately available to direct the cleavage of each probe oligonucleotide that hybridizes to a target nucleic acid. For this reason, in the Examples described herein, the Invader™ oligonucleotide is provided in excess over the probe oligonucleotide; often this excess is 10-fold. While this is an effective ratio, it is not intended that the practice of the present invention be limited to any particular ratio of Invader™-to-probe (a ratio of 2- to 100-fold is contemplated).

Buffer conditions must be chosen that will be compatible with both the oligonucleotide/target hybridization and with the activity of the cleavage agent. The optimal buffer conditions for nucleic acid modification enzymes, and particularly DNA modification enzymes, generally included enough mono- and di-valent salts to allow association of nucleic acid strands by base-pairing. If the method of the present invention is performed using an enzymatic cleavage agent other than those specifically described here, the reactions may generally be performed in any such buffer reported to be optimal for the nuclease function of the cleavage agent. In general, to test the utility of any cleavage agent in this method, test reactions are performed wherein the cleavage agent of interest is tested in the MOPS/$MnCl_2$/KCl buffer or Mg-containing buffers described herein and in whatever buffer has been reported to be suitable for use with that agent, in a manufacturer's data sheet, a journal article, or in personal communication.

The products of the Invader™-directed cleavage reaction are fragments generated by structure-specific cleavage of the input oligonucleotides. The resulting cleaved and/or uncleaved oligonucleotides may be analyzed and resolved by a number of methods including electrophoresis (on a variety of supports including acrylamide or agarose gels, paper, etc.), chromatography, fluorescence polarization, mass spectrometry and chip hybridization. The invention is illustrated using electrophoretic separation for the analysis of the products of the cleavage reactions. However, it is noted that the resolution of the cleavage products is not limited to electrophoresis. Electrophoresis is chosen to illustrate the method of the invention

because electrophoresis is widely practiced in the art and is easily accessible to the average practitioner.

The probe and Invader™ oligonucleotides may contain a label to aid in their detection following the cleavage reaction. The label may be a radioisotope (*e.g.*, a ^{32}P or ^{35}S-labelled nucleotide) placed at either the 5' or 3' end of the oligonucleotide or alternatively, the label may be distributed throughout the oligonucleotide (*i.e.*, a uniformly labelled oligonucleotide). The label may be a nonisotopic detectable moiety, such as a fluorophore, which can be detected directly,or a reactive group which permits specific recognition by a secondary agent. For example, biotinylated oligonucleotides may be detected by probing with a streptavidin molecule which is coupled to an indicator (*e.g.*, alkaline phosphatase or a fluorophore) or a hapten such as digoxigenin may be detected using a specific antibody coupled to a similar indicator.

Optimization Of Reaction Conditions

The Invader™-directed cleavage reaction is useful to detect the presence of specific nucleic acids. In addition to the considerations listed above for the selection and design of the Invader™ and probe oligonucleotides, the conditions under which the reaction is to be performed may be optimized for detection of a specific target sequence.

One objective in optimizing the Invader™-directed cleavage assay is to allow specific detection of the fewest copies of a target nucleic acid. To achieve this end, it is desirable that the combined elements of the reaction interact with the maximum efficiency, so that the rate of the reaction (*e.g.*, the number of cleavage events per minute) is maximized. Elements contributing to the overall efficiency of the reaction include the rate of hybridization, the rate of cleavage, and the efficiency of the release of the cleaved probe.

The rate of cleavage will be a function of the cleavage means chosen, and may be made optimal according to the manufacturer's instructions when using commercial preparations of enzymes or as described in the examples herein. The other elements (rate of hybridization, efficiency of release) depend upon the execution of the reaction, and optimization of these elements is discussed below.

Three elements of the cleavage reaction that significantly affect the rate of nucleic acid hybridization are the concentration of the nucleic acids, the temperature at which the cleavage reaction is performed and the concentration of salts and/or other charge-shielding ions in the reaction solution.

The concentrations at which oligonucleotide probes are used in assays of this type are well known in the art, and are discussed above. One example of a common approach to optimizing an oligonucleotide concentration is to choose a starting amount of oligonucleotide for pilot tests; 0.01 to 2 μM is a concentration range used in many oligonucleotide-based assays. When initial cleavage reactions are performed, the following questions may be asked of the data: Is the reaction performed in the absence of the target nucleic acid substantially free of the cleavage product?; Is the site of cleavage specifically shifted in accordance with the design of the Invader™ oligonucleotide?; Is the specific cleavage product easily detected in the presence of the uncleaved probe (or is the amount of uncut material overwhelming the chosen visualization method)?

A negative answer to any of these questions would suggest that the probe concentration is too high, and that a set of reactions using serial dilutions of the probe should be performed until the appropriate amount is identified. Once identified for a given target nucleic acid in a give sample type (*e.g.*, purified genomic DNA, body fluid extract, lysed bacterial extract), it should not need to be re-optimized. The sample type is important because the complexity of the material present may influence the probe optimum.

Conversely, if the chosen initial probe concentration is too low, the reaction may be slow, due to inefficient hybridization. Tests with increasing quantities of the probe will identify the point at which the concentration exceeds the optimum. Since the hybridization will be facilitated by excess of probe, it is desirable, but not required, that the reaction be performed using probe concentrations just below this point.

The concentration of Invader™ oligonucleotide can be chosen based on the design considerations discussed above. In a preferred embodiment, the Invader™ oligonucleotide is in excess of the probe oligonucleotide. In a particularly preferred embodiment, the Invader™ is approximately 10-fold more abundant than the probe.

Temperature is also an important factor in the hybridization of oligonucleotides. The range of temperature tested will depend in large part, on the design of the oligonucleotides, as discussed above. In a preferred embodiment, the reactions are performed at temperatures slightly below the T_m of the least stable oligonucleotide in the reaction. Melting temperatures for the oligonucleotides and for their component regions (X, Y and Z, Fig. 25), can be estimated through the use of computer software or, for a more rough approximation, by assigning the value of 2°C per A-T basepair, and 4°C per G-C basepair, and taking the sum across an expanse of nucleic acid. The latter method may be used for oligonucleotides of

approximately 10-30 nucleotides in length. Because even computer prediction of the T_m of a nucleic acid is only an approximation, the reaction temperatures chosen for initial tests should bracket the calculated T_m. While optimizations are not limited to this, 5°C increments are convenient test intervals in these optimization assays.

When temperatures are tested, the results can be analyzed for specificity (the first two of the questions listed above) in the same way as for the oligonucleotide concentration determinations. Non-specific cleavage (*i.e.*, cleavage of the probe at many or all positions along its length) would indicate non-specific interactions between the probe and the sample material, and would suggest that a higher temperature should be employed. Conversely, little or no cleavage would suggest that even the intended hybridization is being prevented, and would suggest the use of lower temperatures. By testing several temperatures, it is possible to identify an approximate temperature optimum, at which the rate of specific cleavage of the probe is highest. If the oligonucleotides have been designed as described above, the T_m of the Z-region of the probe oligonucleotide should be below this temperature, so that turnover is assured.

A third determinant of hybridization efficiency is the salt concentration of the reaction. In large part, the choice of solution conditions will depend on the requirements of the cleavage agent, and for reagents obtained commercially, the manufacturer's instructions are a resource for this information. When developing an assay utilizing any particular cleavage agent, the oligonucleotide and temperature optimizations described above should be performed in the buffer conditions best suited to that cleavage agent.

A "no enzyme" control allows the assessment of the stability of the labeled oligonucleotides under particular reaction conditions, or in the presence of the sample to be tested (*i.e.*, in assessing the sample for contaminating nucleases). In this manner, the substrate and oligonucleotides are placed in a tube containing all reaction components, except the enzyme and treated the same as the enzyme-containing reactions. Other controls may also be included. For example, a reaction with all of the components except the target nucleic acid will serve to confirm the dependence of the cleavage on the presence of the target sequence.

Probing For Multiple Alleles

The Invader™-directed cleavage reaction is also useful in the detection and quantification of individual variants or alleles in a mixed sample population. By way of example, such a need exists in the analysis of tumor material for mutations in genes associated with cancers. Biopsy material from a tumor can have a significant complement of normal cells, so it is desirable to detect mutations even when present in fewer than 5% of the copies of the target nucleic acid in a sample. In this case, it is also desirable to measure what fraction of the population carries the mutation. Similar analyses may also be done to examine allelic variation in other gene systems, and it is not intended that the method of the present invention by limited to the analysis of tumors.

As demonstrated below, reactions can be performed under conditions that prevent the cleavage of probes bearing even a single-nucleotide difference mismatch within the region of the target nucleic acid termed "Z" in Fig. 25, but that permit cleavage of a similar probe that is completely complementary to the target in this region. Thus, the assay may be used to quantitate individual variants or alleles within a mixed sample.

The use of multiple, differently labelled probes in such an assay is also contemplated. To assess the representation of different variants or alleles in a sample, one would provide a mixture of probes such that each allele or variant to be detected would have a specific probe (*i.e.*, perfectly matched to the Z region of the target sequence) with a unique label (*e.g.*, no two variant probes with the same label would be used in a single reaction). These probes would be characterized in advance to ensure that under a single set of reaction conditions, they could be made to give the same rate of signal accumulation when mixed with their respective target nucleic acids. Assembly of a cleavage reaction comprising the mixed probe set, a corresponding Invader™ oligonucleotide, the target nucleic acid sample, and the appropriate cleavage agent, along with performance of the cleavage reaction under conditions such that only the matched probes would cleave, would allow independent quantification of each of the species present, and would therefore indicate their relative representation in the target sample.

III. A Comparison Of Invasive Cleavage And Primer-Directed Cleavage

As discussed herein, the terms "invasive" or "Invader™-directed" cleavage specifically denote the use of a first, upstream oligonucleotide, as defined below, to cause specific cleavage at a site within a second, downstream sequence. To effect such a direction of

cleavage to a region within a duplex, it is required that the first and second oligonucleotides overlap in sequence. That is to say, a portion of the upstream oligonucleotide, termed the "Invader™", has significant homology to a portion of the downstream "probe" oligonucleotide, so that these regions would tend to basepair with the same complementary region of the target nucleic acid to be detected. While not limiting the present invention to any particular mechanism, the overlapping regions would be expected to alternate in their occupation of the shared hybridization site. When the probe oligonucleotide fully anneals to the target nucleic acid, and thus forces the 3' region of the Invader™ to remain unpaired, the structure so formed is not a substrate for the 5' nucleases of the present invention. By contrast, when the inverse is true, the structure so formed is substrate for these enzymes, allowing cleavage and release of the portion of the probe oligonucleotide that is displaced by the Invader™ oligonucleotide. The shifting of the cleavage site to a region the probe oligonucleotide that would otherwise be basepaired to the target sequence is one hallmark of the invasive cleavage assay (*i.e.*, the Invader™-directed cleavage assay) of the present invention.

It is beneficial at this point to contrast the invasive cleavage as described above with two other forms of probe cleavage that may lead to internal cleavage of a probe oligonucleotide, but which do not comprise invasive cleavage. In the first case, a hybridized probe may be subject to duplex-dependent 5' to 3' exonuclease "nibbling," such that the oligonucleotide is shortened from the 5' end until it cannot remain bound to the target (*see, e.g.*, Examples 5-7 and Figs. 26-28). The site at which such nibbling stops can appear to be discrete, and, depending on the difference between the melting temperature of the full-length probe and the temperature of the reaction, this stopping point may be 1 or several nucleotides into the probe oligonucleotide sequence. Such "nibbling" is often indicated by the presence of a "ladder" of longer products ascending size up to that of the full length of the probe, but this is not always the case. While any one of the products of such a nibbling reaction may be made to match in size and cleavage site the products of an invasive cleavage reaction, the creation of these nibbling products would be highly dependent on the temperature of the reaction and the nature of the cleavage agent, but would be independent of the action of an upstream oligonucleotide, and thus could not be construed to involve invasive cleavage.

A second cleavage structure that may be considered is one in which a probe oligonucleotide has several regions of complementarity with the target nucleic acid, interspersed with one or more regions or nucleotides of noncomplementarity. These

noncomplementary regions may be thought of as "bubbles" within the nucleic acid duplex. As temperature is elevated, the regions of complementarity can be expected to "melt" in the order of their stability, lowest to highest. When a region of lower stability is near the end of a segment of duplex, and the next region of complementarity along the strand has a higher melting temperature, a temperature can be found that will cause the terminal region of duplex to melt first, opening the first bubble, and thereby creating a preferred substrate structure of the cleavage by the 5' nucleases of the present invention (Fig. 36A). The site of such cleavage would be expected to be on the 5' arm, within 2 nucleotides of the junction between the single and double-stranded regions (Lyamichev *et al.*, *supra.* and U.S. Patent No. 5,422,253)

An additional oligonucleotide could be introduced to basepair along the target nucleic acid would have a similar effect of opening this bubble for subsequent cleavage of the unpaired 5' arm (Fig. 36B and Fig. 6). Note in this case, the 3' terminal nucleotides of the upstream oligonucleotide anneals along the target nucleic acid sequence in such a manner that the 3' end is located within the "bubble" region. Depending on the precise location of the 3' end of this oligonucleotide, the cleavage site may be along the newly unpaired 5' arm, or at the site expected for the thermally opened bubble structure as described above. In the former case the cleavage is not within a duplexed region, and is thus not invasive cleavage, while in the latter the oligonucleotide is merely an aide in inducing cleavage at a site that might otherwise be exposed through the use of temperature alone (*i.e.*, in the absence of the additional oligonucleotide), and is thus not considered to be invasive cleavage.

In summary, any arrangement of oligonucleotides used for the cleavage-based detection of a target sequence can be analyzed to determine if the arrangement is an invasive cleavage structure as contemplated herein. An invasive cleavage structure supports cleavage of the probe in a region that, in the absence of an upstream oligonucleotide, would be expected to be basepaired to the target nucleic acid.

Example 26 below provides further guidance for the design and execution of a experiments which allow the determination of whether a given arrangement of a pair of upstream and downstream (*i.e.*, the probe) oligonucleotides when annealed along a target nucleic acid would form an invasive cleavage structure.

IV. Fractionation Of Specific Nucleic Acids By Selective Charge Reversal

Some nucleic acid-based detection assays involve the elongation and/or shortening of oligonucleotide probes. For example, as described herein, the primer-directed, primer-independent, and Invader™-directed cleavage assays, as well as the "nibbling" assay all involve the cleavage (*i.e.*, shortening) of oligonucleotides as a means for detecting the presence of a target nucleic sequence. Examples of other detection assays which involve the shortening of an oligonucleotide probe include the "TaqMan" or nick-translation PCR assay described in U.S. Patent No. 5,210,015 to Gelfand *et al.* (the disclosure of which is herein incorporated by reference), the assays described in U.S. Patent Nos. 4,775,619 and 5,118,605 to Urdea (the disclosures of which are herein incorporated by reference), the catalytic hybridization amplification assay described in U.S. Patent No. 5,403,711 to Walder and Walder (the disclosure of which is herein incorporated by reference), and the cycling probe assay described in U.S. Patents Nos. 4,876,187 and 5,011,769 to Duck *et al.* (the disclosures of which are herein incorporated by reference). Examples of detection assays which involve the elongation of an oligonucleotide probe (or primer) include the polymerase chain reaction (PCR) described in U.S. Patent Nos. 4,683,195 and 4,683,202 to Mullis and Mullis *et al.* (the disclosures of which are herein incorporated by reference) and the ligase chain reaction (LCR) described in U.S. Patent Nos. 5,427,930 and 5,494,810 to Birkenmeyer *et al.* and Barany *et al.* (the disclosures of which are herein incorporated by reference). The above examples are intended to be illustrative of nucleic acid-based detection assays that involve the elongation and/or shortening of oligonucleotide probes and do not provide an exhaustive list.

Typically, nucleic acid-based detection assays that involve the elongation and/or shortening of oligonucleotide probes require post-reaction analysis to detect the products of the reaction. It is common that, the specific reaction product(s) must be separated from the other reaction components, including the input or unreacted oligonucleotide probe. One detection technique involves the electrophoretic separation of the reacted and unreacted oligonucleotide probe. When the assay involves the cleavage or shortening of the probe, the unreacted product will be longer than the reacted or cleaved product. When the assay involves the elongation of the probe (or primer), the reaction products will be greater in length than the input. Gel-based electrophoresis of a sample containing nucleic acid molecules of different lengths separates these fragments primarily on the basis of size. This is due to the fact that in solutions having a neutral or alkaline pH, nucleic acids having

widely different sizes (*i.e.*, molecular weights) possess very similar charge-to-mass ratios and do not separate (Andrews, Electrophoresis, 2nd Edition, Oxford University Press (1986), pp. 153-154]. The gel matrix acts as a molecular sieve and allows nucleic acids to be separated on the basis of size and shape (*e.g.*, linear, relaxed circular or covalently closed supercoiled circles).

Unmodified nucleic acids have a net negative charge due to the presence of negatively charged phosphate groups contained within the sugar-phosphate backbone of the nucleic acid. Typically, the sample is applied to gel near the negative pole and the nucleic acid fragments migrate into the gel toward the positive pole with the smallest fragments moving fastest through the gel.

The present invention provides a novel means for fractionating nucleic acid fragments on the basis of charge. This novel separation technique is related to the observation that positively charged adducts can affect the electrophoretic behavior of small oligonucleotides because the charge of the adduct is significant relative to charge of the whole complex. In addition, to the use of positively charged adducts (*e.g.*, Cy3 and Cy5 fluorescent dyes, the positively charged heterodimeric DNA-binding dyes shown in Fig. 66, etc.), the oligonucleotide may contain amino acids (particularly useful amino acids are the charged amino acids: lysine, arginine, asparate, glutamate), modified bases, such as amino-modified bases, and/or a phosphonate backbone (at all or a subset of the positions). In addition as discussed further below, a neutral dye or detection moiety (*e.g.*, biotin, streptavidin, etc.) may be employed in place of a positively charged adduct in conjunction with the use of amino-modified bases and/or a complete or partial phosphonate backbone.

This observed effect is of particular utility in assays based on the cleavage of DNA molecules. Using the assays described herein as an example, when an oligonucleotide is shortened through the action of a Cleavase® enzyme or other cleavage agent, the positive charge can be made to not only significantly reduce the net negative charge, but to actually override it, effectively "flipping" the net charge of the labeled entity. This reversal of charge allows the products of target-specific cleavage to be partitioned from uncleaved probe by extremely simple means. For example, the products of cleavage can be made to migrate towards a negative electrode placed at any point in a reaction vessel, for focused detection without gel-based electrophoresis; Example 24 provides examples of devices suitable for focused detection without gel-based electrophoresis. When a slab gel is used, sample wells can be positioned in the center of the gel, so that the cleaved and uncleaved probes can be

observed to migrate in opposite directions. Alternatively, a traditional vertical gel can be used, but with the electrodes reversed relative to usual DNA gels (*i.e.*, the positive electrode at the top and the negative electrode at the bottom) so that the cleaved molecules enter the gel, while the uncleaved disperse into the upper reservoir of electrophoresis buffer.

An important benefit of this type of readout is the absolute nature of the partition of products from substrates (*i.e.*, the separation is virtually 100%). This means that an abundance of uncleaved probe can be supplied to drive the hybridization step of the probe-based assay, yet the unconsumed (*i.e.*, unreacted) probe can, in essence, be subtracted from the result to reduce background by virtue of the fact that the unreacted probe will not migrate to the same pole as the specific reaction product.

Through the use of multiple positively charged adducts, synthetic molecules can be constructed with sufficient modification that the normally negatively charged strand is made nearly neutral. When so constructed, the presence or absence of a single phosphate group can mean the difference between a net negative or a net positive charge. This observation has particular utility when one objective is to discriminate between enzymatically generated fragments of DNA, which lack a 3' phosphate, and the products of thermal degradation, which retain a 3' phosphate (and thus two additional negative charges). Examples 22 and 23 demonstrate the ability to separate positively charged reaction products from a net negatively charged substrate oligonucleotide. As discussed in these examples, oligonucleotides may be transformed from net negative to net positively charged compounds. In Example 23, the positively charged dye, Cy3 was incorporated at the 5' end of a 22-mer (SEQ ID NO:50) which also contained two amino-substituted residues at the 5' end of the oligonucleotide; this oligonucleotide probe carries a net negative charge. After cleavage, which occurred 2 nucleotides into the probe, the following labelled oligonucleotide was released: 5'-Cy3-AminoT-AminoT-3' (as well as the remaining 20 nucleotides of SEQ ID NO:50). This short fragment bears a net positive charge while the remainder of the cleaved oligonucleotide and the unreacted or input oligonucleotide bear net negative charges.

The present invention contemplates embodiments wherein the specific reaction product produced by any cleavage of any oligonucleotide can be designed to carry a net positive charge while the unreacted probe is charge neutral or carries a net negative charge. The present invention also contemplates embodiments where the released product may be designed to carry a net negative charge while the input nucleic acid carries a net positive charge. Depending on the length of the released product to be detected, positively charged dyes may

be incorporated at the one end of the probe and modified bases may be placed along the oligonucleotide such that upon cleavage, the released fragment containing the positively charged dye carries a net positive charge. Amino-modified bases may be used to balance the charge of the released fragment in cases where the presence of the positively charged adduct (*e.g.*, dye) alone is not sufficient to impart a net positive charge on the released fragment. In addition, the phosphate backbone may be replaced with a phosphonate backbone at a level sufficient to impart a net positive charge (this is particularly useful when the sequence of the oligonucleotide is not amenable to the use of amino-substituted bases); Figs. 45 and 46 show the structure of short oligonucleotides containing a phosphonate group on the second T residue). An oligonucleotide containing a fully phosphonate-substituted backbone would be charge neutral (absent the presence of modified charged residues bearing a charge or the presence of a charged adduct) due to the absence of the negatively charged phosphate groups. Phosphonate-containing nucleotides (*e.g.*, methylphosphonate-containing nucleotides are readily available and can be incorporated at any position of an oligonucleotide during synthesis using techniques which are well known in the art.

In essence, the invention contemplates the use of charge-based separation to permit the separation of specific reaction products from the input oligonucleotides in nucleic acid-based detection assays. The foundation of this novel separation technique is the design and use of oligonucleotide probes (typically termed "primers" in the case of PCR) which are "charge balanced" so that upon either cleavage or elongation of the probe it becomes "charge unbalanced," and the specific reaction products may be separated from the input reactants on the basis of the net charge.

In the context of assays which involve the elongation of an oligonucleotide probe (*i.e.*, a primer), such as is the case in PCR, the input primers are designed to carry a net positive charge. Elongation of the short oligonucleotide primer during polymerization will generate PCR products which now carry a net negative charge. The specific reaction products may then easily be separated and concentrated away from the input primers using the charge-based separation technique described herein (the electrodes will be reversed relative to the description in Example 23 as the product to be separated and concentrated after a PCR will carry a negative charge).

V. Invader™-Directed Cleavage Using Miniprobes And Mid-Range Probes

As discussed in Section III above, the Invader™-directed cleavage assay may be performed using Invader™ and probe oligonucleotides which have a length of about 13-25 nucleotides (typically 20-25 nucleotides). It is also contemplated that the oligonucleotides that span the X, Y and Z regions (see Fig. 25), the Invader™ and probe oligonucleotides, may themselves be composed of shorter oligonucleotide sequences that align along a target strand but that are not covalently linked. This is to say that there is a nick in the sugar-phosphate backbone of the composite oligonucleotide, but that there is no disruption in the progression of base-paired nucleotides in the resulting duplex. When short strands of nucleic acid align contiguously along a longer strand the hybridization of each is stabilized by the hybridization of the neighboring fragments because the basepairs can stack along the helix as though the backbone was in fact uninterrupted. This cooperativity of binding can give each segment a stability of interaction in excess of what would be expected for the segment hybridizing to the longer nucleic acid alone. One application of this observation has been to assemble primers for DNA sequencing, typically about 18 nucleotides long, from sets of three hexamer oligonucleotides that are designed to hybridize in this way (Kotler *et al.* Proc. Natl. Acad. Sci. USA 90:4241 [1993]). The resulting doubly-nicked primer can be extended enzymatically in reactions performed at temperatures that might be expected to disrupt the hybridization of hexamers, but not of 18-mers.

The use of composite or split oligonuceotides is applied with success in the Invader™-directed cleavage assay. The probe oligonucleotide may be split into two oligonucleotides which anneal in a contiguous and adjacent manner along a target oligonucleotide as diagrammed in Fig. 57. In this Figure, the downstream oligonucleotide (analogous to the probe of Fig. 25) is assembled from two smaller pieces: a short segment of 6-10 nts (termed the "miniprobe"), that is to be cleaved in the course of the detection reaction, and an oligonucleotide that hybridizes immediately downstream of the miniprobe (termed the "stacker"), which serves to stabilize the hybridization of the probe. To form the cleavage structure, an upstream oligonucleotide (the "Invader™" oligo) is provided to direct the cleavage activity to the desired region of the miniprobe. Assembly of the probe from non-linked pieces of nucleic acid (*i.e.*, the miniprobe and the stacker) allows regions of sequences to be changed without requiring the re-synthesis of the entire proven sequence, thus improving the cost and flexibility of the detection system. In addition, the use of unlinked

composite oligonucleotides makes the system more stringent in its requirement of perfectly matched hybridization to achieve signal generation, allowing this to be used as a sensitive means of detecting mutations or changes in the target nucleic acid sequences.

As illustrated in Fig. 57, in one embodiment, the methods of the present invention employ at least three oligonucleotides that interact with a target nucleic acid to form a cleavage structure for a structure-specific nuclease. More specifically, the cleavage structure comprises i) a target nucleic acid that may be either single-stranded or double-stranded (when a double-stranded target nucleic acid is employed, it may be rendered single-stranded, *e.g.*, by heating); ii) a first oligonucleotide, termed the "stacker," which defines a first region of the target nucleic acid sequence by being the complement of that region (region W of the target as shown in Fig. 57); iii) a second oligonucleotide, termed the "miniprobe," which defines a second region of the target nucleic acid sequence by being the complement of that region (regions X and Z of the target as shown in Fig. 57); iv) a third oligonucleotide, termed the "Invader™," the 5' part of which defines a third region of the same target nucleic acid sequence (regions Y and X in Fig. 57), adjacent to and downstream of the second target region (regions X and Z), and the second or 3' part of which overlaps into the region defined by the second oligonucleotide (region X depicts the region of overlap). The resulting structure is diagrammed in Fig. 57.

While not limiting the invention or the instant discussion to any particular mechanism of action, the diagram in Fig. 57 represents the effect on the site of cleavage caused by this type of arrangement of three oligonucleotides. The design of these three oligonucleotides is described below in detail. In Fig. 57, the 3' ends of the nucleic acids (*i.e.*, the target and the oligonucleotides) are indicated by the use of the arrowheads on the ends of the lines depicting the strands of the nucleic acids (and where space permits, these ends are also labelled "3'"). It is readily appreciated that the three oligonucleotides (the Invader™, the miniprobe and the stacker) are arranged in a parallel orientation relative to one another, while the target nucleic acid strand is arranged in an anti-parallel orientation relative to the three oligonucleotides. Further it is clear that the Invader™ oligonucleotide is located upstream of the miniprobe oligonucleotide and that the miniprobe olignuceotide is located upstream of the stacker oligonucleotide and that with respect to the target nucleic acid strand, region W is upstream of region Z, region Z is upstream of upstream of region X and region X is upstream of region Y (that is region Y is downstream of region X, region X is downstream of region Z and region Z is downstream of region W). Regions of complementarity between the opposing

strands are indicated by the short vertical lines. While not intended to indicate the precise location of the site(s) of cleavage, the area to which the site of cleavage within the miniprobe oligonucleotide is shifted by the presence of the Invader™ oligonucleotide is indicated by the solid vertical arrowhead. Fig. 57 is not intended to represent the actual mechanism of action or physical arrangement of the cleavage structure and further it is not intended that the method of the present invention be limited to any particular mechanism of action.

It can be considered that the binding of these oligonucleotides divides the target nucleic acid into four distinct regions: one region that has complementarity to only the stacker (shown as "W"); one region that has complementarity to only the miniprobe (shown as "Z"); one region that has complementarity only to the Invader™ oligo (shown as "Y"); and one region that has complementarity to both the Invader™ and miniprobe oligonucleotides (shown as "X").

In addition to the benefits cited above, the use of a composite design for the oligonucleotides which form the cleavage structure allows more latitude in the design of the reaction conditions for performing the Invader™-directed cleavage assay. When a longer probe (*e.g.*, 16-25 nt), as described in Section III above, is used for detection in reactions that are performed at temperatures below the T_m of that probe, the cleavage of the probe may play a significant role in destabilizing the duplex of which it is a part, thus allowing turnover and reuse of the recognition site on the target nucleic acid. In contrast, with miniprobes, reaction temperatures that are at or above the T_m of the probe mean that the probe molecules are hybridizing and releasing from the target quite rapidly even without cleavage of the probe. When an upstream Invader™ oligonucleotide and a cleavage means are provided the miniprobe will be specifically cleaved, but the cleavage will not be necessary to the turnover of the miniprobe. If a long probe (*e.g.*, 16-25 nt) were to be used in this way the temperatures required to achieve this state would be quite high, around 65 to 70°C for a 25-mer of average base composition. Requiring the use of such elevated temperatures limits the choice of cleavage agents to those that are very thermostable, and may contribute to background in the reactions, depending of the means of detection, through thermal degradation of the probe oligonucleotides. Thus, the shorter probes are preferable for use in this way.

The miniprobe of the present invention may vary in size depending on the desired application. In one embodiment, the probe may be relatively short compared to a standard probe (*e.g.*, 16-25 nt), in the range of 6 to 10 nucleotides. When such a short probe is used

reaction conditions can be chosen that prevent hybridization of the miniprobe in the absence of the stacker oligonucleotide. In this way a short probe can be made to assume the statistical specificity and selectivity of a longer sequence. In the event of a perturbation in the cooperative binding of the miniprobe and stacker nucleic acids, as might be caused by a mismatch within the short sequence (*i.e.*, region "Z" which is the region of the miniprobe which does not overlap with the Invader™) or at the junction between the contiguous duplexes, this cooperativity can be lost, dramatically reducing the stability of the shorter oligonucleotide (*i.e.*, the miniprobe), and thus reducing the level of cleaved product in the assay of the present invention.

It is also contemplated that probes of intermediate size may be used. Such probes, in the 11 to 15 nucleotide range, may blend some of the features associated with the longer probes as originally described, these features including the ability to hybridize and be cleaved absent the help of a stacker oligonucleotide. At temperatures below the expected T_m of such probes, the mechanisms of turnover may be as discussed above for probes in the 20 nt range, and be dependent on the removal of the sequence in the 'X' region for destabilization and cycling.

The mid-range probes may also be used at elevated temperatures, at or above their expected T_m, to allow melting rather than cleavage to promote probe turnover. In contrast to the longer probes described above, however, the temperatures required to allow the use of such a thermally driven turnover are much lower (about 40 to 60°C), thus preserving both the cleavage means and the nucleic acids in the reaction from thermal degradation. In this way, the mid-range probes may perform in some instances like the miniprobes described above. In a further similarity to the miniprobes, the accumulation of cleavage signal from a mid-range probe may be helped under some reaction conditions by the presence of a stacker.

To summarize, a standard long probe usually does not benefit from the presence of a stacker oligonucleotide downstream (the exception being cases where such an oligonucleotide may also disrupt structures in the target nucleic acid that interfere with the probe binding), and it is usually used in conditions requiring several nucleotides to be removed to allow the oligonucleotide to release from the target efficiently.

The miniprobe is very short and performs optimally in the presence of a downstream stacker oligonucleotide. The miniprobes are well suited to reactions conditions that use the temperature of the reaction to drive rapid exchange of the probes on the target regardless of

whether any bases have been cleaved. In reactions with sufficient amount of the cleavage means, the probes that do bind will be rapidly cleaved before they melt off.

The mid-range or midiprobe combines features of these probes and can be used in reactions like those designed long probes, with longer regions of overlap ("X" regions) to drive probe turnover at lower temperature. In a preferred embodiment, the midrange probes are used at temperatures sufficiently high that the probes are hybridizing to the target and releasing rapidly regardless of cleavage. This is known to be the behavior of oligonucleotides at or near their melting temperature. This mode of turnover is more similar to that used with miniprobe/stacker combinations than with long probes. The mid-range probe may have enhanced performance in the presence of a stacker under some circumstances. For example, with a probe in the lower end of the mid-range (*e.g.*, 11 nt), or one with exceptional A/T content, in a reaction performed well in excess of the T_m of the probe (*e.g.*, >10°C above) the presence of a stacker would be likely to enhance the performance of the probe, while at a more moderate temperature the probe may be indifferent to a stacker.

The distinctions between the mini-, midi- (*i.e.*, mid-range) and long probes are not contemplated to be inflexible and based only on length. The performance of any given probe may vary with its specific sequence, the choice of solution conditions, the choice of temperature and the selected cleavage means.

It is shown in Example 18 that the assemblage of oligonucleotides that comprises the cleavage structure of the present invention is sensitive to mismatches between the probe and the target. The site of the mismatch used in Ex. 18 provides one example and is not intended to be a limitation in location of a mismatch affecting cleavage. It is also contemplated that a mismatch between the Invader™ oligonucleotide and the target may be used to distinguish related target sequences. In the 3-oligonucleotide system, comprising an Invader™, a probe and a stacker oligonucleotide, it is contemplated that mismatches may be located within any of the regions of duplex formed between these oligonucleotides and the target sequence. In a preferred embodiment, a mismatch to be detected is located in the probe. In a particularly preferred embodiment, the mismatch is in the probe, at the basepair immediately upstream (*i.e.*, 5') of the site that is cleaved when the probe is not mismatched to the target.

In another preferred embodiment, a mismatch to be detected is located within the region 'Z' defined by the hybridization of a miniprobe. In a particularly preferred embodiment, the mismatch is in the miniprobe, at the basepair immediately upstream (*i.e.*, 5') of the site that is cleaved when the miniprobe is not mismatched to the target.

It is also contemplated that different sequences may be detected in a single reaction. Probes specific for the different sequences may be differently labeled. For example, the probes may have different dyes or other detectable moieties, different lengths, or they may have differences in net charges of the products after cleavage. When differently labeled in one of these ways, the contribution of each specific target sequence to final product can be tallied. This has application in detecting the quantities of different versions of a gene within a mixture. Different genes in a mixture to be detected and quantified may be wild type and mutant genes (*e.g.*, as may be found in a tumor sample, such as a biopsy). In this embodiment, one might design the probes to precisely the same site, but one to match the wild-type sequence and one to match the mutant. Quantitative detection of the products of cleavage from a reaction performed for a set amount of time will reveal the ratio of the two genes in the mixture. Such analysis may also be performed on unrelated genes in a mixture. This type of analysis is not intended to be limited to two genes. Many variants within a mixture may be similarly measured.

Alternatively, different sites on a single gene may be monitored and quantified to verify the measurement of that gene. In this embodiment, the signal from each probe would be expected to be the same.

It is also contemplated that multiple probes may be used that are not differently labeled, such that the aggregate signal is measured. This may be desirable when using many probes designed to detect a single gene to boost the signal from that gene. This configuration may also be used for detecting unrelated sequences within a mix. For example, in blood banking it is desirable to know if any one of a host of infectious agents is present in a sample of blood. Because the blood is discarded regardless of which agent is present, different signals on the probes would not be required in such an application of the present invention, and may actually be undesirable for reasons of confidentiality.

Just as described for the two-oligonucleotide system, above, the specificity of the detection reaction will be influenced by the aggregate length of the target nucleic acid sequences involved in the hybridization of the complete set of the detection oligonucleotides. For example, there may be applications in which it is desirable to detect a single region within a complex genome. In such a case the set of oligonucleotides may be chosen to require accurate recognition by hybridization of a longer segment of a target nucleic acid, often in the range of 20 to 40 nucleotides. In other instances it may be desirable to have the set of oligonucleotides interact with multiple sites within a target sample. In these cases one

approach would be to use a set of oligonucleotides that recognize a smaller, and thus statistically more common, segment of target nucleic acid sequence.

In one preferred embodiment, the Invader™ and stacker oligonucleotides may be designed to be maximally stable, so that they will remain bound to the target sequence for extended periods during the reaction. This may be accomplished through any one of a number of measures well known to those skilled in the art, such as adding extra hybridizing sequences to the length of the oligonucleotide (up to about 50 nts in total length), or by using residues with reduced negative charge, such as phosphorothioates or peptide-nucleic acid residues, so that the complementary strands do not repel each other to degree that natural strands do. Such modifications may also serve to make these flanking oligonucleotides resistant to contaminating nucleases, thus further ensuring their continued presence on the target strand during the course of the reaction. In addition, the Invader™ and stacker oligonucleotides may be covalently attached to the target (*e.g.*, through the use of psoralen cross-linking).

The use of the reaction temperatures at or near the T_m of the probe oligonucleotide, rather than that used for cleavage, to drive the turnover of the probe oligonucleotide in these detection reactions means that the amount of the probe oligonucleotide cleaved off may be substantially reduced without adversely affecting the turnover rate. It has been determined that the relationship between the 3' end of the upstream oligonucleotide and the desired site of cleavage on the probe must be carefully designed. It is known that the preferred site of cleavage for the types of structure specific endonucleases employed herein is one basepair into a duplex (Lyamichev *et al., supra*). It was previously believed that the presence of an upstream oligonucleotide or primer allowed the cleavage site to be shifted away from this preferred site, into the single stranded region of the 5' arm (Lyamichev *et al., supra* and U.S. Patent No. 5,422,253). In contrast to this previously proposed mechanism, and while not limiting the present invention to any particular mechanism, it is believed that the nucleotide immediately 5', or upstream of the cleavage site on the probe (including miniprobe and mid-range probes) must be able to basepair with the target for efficient cleavage to occur. In the case of the present invention, this would be the nucleotide in the probe sequence immediately upstream of the intended cleavage site. In addition, as described herein, it has been observed that in order to direct cleavage to that same site in the probe, the upstream oligonucleotide must have its 3' base (*i.e.*, nt) immediately upstream of the intended cleavage site of the probe. This places the 3' terminal nucleotide of the upstream oligonucleotide and the base of

the probe oligonucleotide 5' of the cleavage site in competition for pairing with the corresponding nucleotide of the target strand.

To examine the outcome of this competition (*i.e.* which base is paired during a successful cleavage event), substitutions were made in the probe and Invader™ oligonucleotides such that either the probe or the Invader™ oligonucleotide were mismatched with the target sequence at this position. The effects of both arrangements on the rates of cleavage were examined. When the Invader™ oligonucleotide is unpaired at the 3' end, the rate of cleavage was not reduced. If this base was removed, however, the cleavage site was shifted upstream of the intended site. In contrast, if the probe oligonucleotide was not base-paired to the target just upstream of the site to which the Invader™ oligonucleotide was directing cleavage, the rate of cleavage was dramatically reduced, suggesting that when a competition exists, the probe oligonucleotide was the molecule to be base-paired in this position.

It appears that the 3' end of the upstream Invader™ oligonucleotide is unpaired during cleavage, and yet is required for accurate positioning of the cleavage. To examine which part(s) of the 3' terminal nucleotide are required for the positioning of cleavage, Invader™ oligonucleotides were designed that terminated on this end with nucleotides that were altered in a variety of ways. Sugars examined included 2' deoxyribose with a 3' phosphate group, a dideoxyribose, 3' deoxyribose, 2' O-methyl ribose, arabinose and arabinose with a 3' phosphate. Abasic ribose, with and without 3' phosphate were tested. Synthetic "universal" bases such at 3-nitropyrrole and 5-3nitroindole on ribose sugars were tested. Finally, a base-like aromatic ring structure, acridine, linked to the 3' end the previous nucleotide without a sugar group was tested. The results obtained support the conclusion that the aromatic ring of the base (at the 3' end of the Invader™ oligonuceotide) is the required moiety for accomplishing the direction of cleavage to the desired site within the downstream probe.

VI. Signal Enhancement By Tailing Of Reaction Products In The Invader™-Directed Cleavage Assay

It has been determined that when oligonucleotide probes are used in cleavage detection assays at elevated temperature, some fraction of the truncated probes will have been shortened by nonspecific thermal degradation, and that such breakage products can make the analysis of the target-specific cleavage data more difficult. The thermal degradation that creates a background ladder of bands when the probes of the present invention are treated at high

temperature for more than a few minutes occurs as a two step process. In the first step the N-glycosyl bond breaks, leaving an abasic site in the DNA strand. At the abasic site the DNA chain is weakened and undergoes spontaneous cleavage through a beta-elimination process. It has been determined that purine bases are about 20 times more prone to breakage than pyrimidine bases (Lindahl, Nature 362:709 [1993]). This suggests that one way of reducing background in methods using oligonucleotides at elevated temperatures is to select target sequences that allow the use of pyrimidine-rich probes. It is preferable, where possible, to use oligonucleotides that are entirely composed of pyrimidine residues. If only one or a few purines are used, the background breakage will appear primarily at the corresponding sites, and these bands (due to thermal breakdown) may be mistaken for the intended cleavage products if care is not taken in the data analysis (*i.e.*, proper controls must be run).

Background cleavage due to thermal breakdown of probe oligonucleotides can, when not resolved from specific cleavage products, reduce the accuracy of quantitation of target nucleic acids based on the amount of accumulated product in a set timeframe. One means of distinguishing the specific from the nonspecific products is disclosed above, and is based on partitioning the products of these reactions by differences in the net charges carried by the different molecular species in the reaction. As was noted in that discussion, the thermal breakage products usually retain 3' phosphates after breakage, while the enzyme-cleaved products do not. The two negative charges on the phosphate facilitate charge-based partition of the products.

The absence of a 3' phosphate on the desired subset of the probe fragments may be used to advantage in enzymatic assays as well. Nucleic acid polymerases, both non-templated (*e.g.*, terminal deoxynucleotidyl transferase, polyA polymerase) and template-dependent (*e.g.*, Pol I-type DNA polymerases), require an available 3' hydroxyl by which to attach further nucleotides. This enzymatic selection of 3' end structure may be used as an effective means of partitioning specific from non-specific products.

In addition to the benefits of the partitioning described above, the addition of nucleotides to the end of the specific product of an Invader™-specific cleavage offers an opportunity to either add label to the products, to add capturable tails to facilitate solid-support based readout systems, or to do both of these things at the same time. Some possible embodiments of this concept are illustrated in Fig. 56.

In Fig. 56, an Invader™ cleavage struture comprising an Invader™ oligonuclotide containing a blocked or non-extendible 3' end (*e.g.*, a 3' dideoxynucleotide) and a probe

oligonucleotide containing a blocked or non-extendable 3' end (the open circle at the 3' end of the oligonucleotides represents a non-extendible nucleotide) and a target nucleic acid is shown; the probe oligonucleotide may contain a 5' end label such as a biotin or a fluorescein (indicated by the stars) label (cleavage structures which employ a 5' biotin-labeled probe or a 5' fluorescein-labeled probe are shown below the large diagram of the cleavage structure to the left and the right, respectively). Following, cleavage of the probe (the site of cleavage is indicated by the large arrowhead), the cleaved biotin-labeled probe is extended using a template-independent polymerase (*e.g.*, TdT) and fluoresceinated nucleotide triphosphates. The fluorescein tailed cleaved probe molecule is then captured by binding via its 5' biotin label to streptavidin and the fluroescence is then measured. Alternatively, following, cleavage of a 5'-fluoresceinated probe, the cleaved probe is extended using a template-independent polymerase (*e.g.*, TdT) and dATP. The polyadenylated (A-tailed) cleaved probe molecule is then captured by binding via the polyA tail to oligo dT attached to a solid support.

The examples described in Fig. 56 are based on the use of TdT to tail the specific products of Invader™-directed cleavage. The description of the use of this particular enzyme is presented by way of example and is not intended as a limitation (indeed, when probe oligos comprising RNA are employed, cleaved RNA probes may be extended using polyA polymerase). It is contemplated that an assay of this type could be configured to use a template-dependent polymerase, as described above. While this would require the presence of a suitable copy template distinct from the target nucleic acid, on which the truncated oligonucleotide could prime synthesis, it can be envisaged that a probe which before cleavage would be unextendible, due to either mismatch or modification of the 3' end, could be activated as a primer when cleaved by an Invader™ directed cleavage. A template directed tailing reaction also has the advantage of allowing greater selection and control of the nucleotides incorporated.

The use of nontemplated tailing does not require the presence of any additional nucleic acids in the detection reaction, avoiding one step of assay development and troubleshooting. In addition, the use of non templated synthesis eliminated the step of hybridization, potentially speeding up the assay. Furthermore, the TdT enzyme is fast, able to add at least >700 nucleotides to substrate oligonucleotides in a 15 minute reaction.

As mentioned above, the tails added can be used in a number of ways. It can be used as a straight-forward way of adding labeled moieties to the cleavage product to increase signal from each cleavage event. Such a reaction is depicted in the left side of Fig. 66. The

labeled moieties may be anything that can, when attached to a nucleotide, be added by the tailing enzyme, such as dye molecules, haptens such as digoxigenin, or other binding groups such as biotin.

In a preferred embodiment the assay includes a means of specifically capturing or partitioning the tailed Invader™-directed cleavage products in the mixture. It can be seen that target nucleic acids in the mixture may be tailed during the reaction. If a label is added, it is desirable to partition the tailed Invader™-directed cleavage products from these other labeled molecules to avoid background in the results. This is easily done if only the cleavage product is capable of being captured. For example, consider a cleavage assay of the present invention in which the probe used has a biotin on the 5' end and is blocked from extension on the 3' end, and in which a dye is added during tailing. Consider further that the products are to be captured onto a support via the biotin moeity, and the captured dye measured to assess the presence of the target nucleic acid. When the label is added by tailing, only the specifically cleaved probes will be labeled. The residual uncut probes can still bind in the final capture step, but they will not contribute to the signal. In the same reaction, nicks and cuts in the target nucleic acid may be tailed by the enzyme, and thus become dye labeled. In the final capture these labeled targets will not bind to the support and thus, though labeled, they will not contribute to the signal. If the final specific product is considered to consist of two portions, the probe-derived portion and the tail portion, can be seen from this discussion that it is particularly preferred that when the probe-derived portion is used for specific capture, whether by hybridization, biotin/streptavidin, or other method, that the label be associated with the tail portion. Conversely, if a label is attached to the probe-derived portion, then the tail portion may be made suitable for capture, as depicted on the right side of Fig. 66. Tails may be captured in a number of ways, including hybridization, biotin incorporation with streptavidin capture, or by virtue if the fact that the longer molecules bind more predictably and efficiently to a number of nucleic acid minding matrices, such as nitrocellulose, nylon, or glass, in membrane, paper, resin, or other form. While not required for this assay, this separation of functions allows effective exclusion from signal of both unreacted probe and tailed target nucleic acid.

In addition to the supports decribed above, the tailed products may be captured onto any support that contains a suitable capture moiety. For example, biotinylated products are generally captured with avidin-treated surfaces. These avidin surfaces may be in microtitre plate wells, on beads, on dipsticks, to name just a few of the possibilities. Such surfaces can

also be modified to contain specific oligonucleotides, allowing capture of product by hybridization. Capture surfaces as described here are generally known to those skilled in the art and include nitrocellulose dipsticks (*e.g.*, GeneComb™, BioRad, Hercules, CA).

VII. Improved Enzymes For Use In Invader™-Directed Cleavage Reactions

A cleavage structure is defined herein as a structure which is formed by the interaction of a probe oligonucleotide and a target nucleic acid to form a duplex, the resulting structure being cleavable by a cleavage means, including but not limited to an enzyme. The cleavage structure is further defined as a substrate for specific cleavage by the cleavage means in contrast to a nucleic acid molecule which is a substrate for nonspecific cleavage by agents such as phosphodiesterases. Examples of some possible cleavage structures are shown in Fig. 15. In considering improvements to enzymatic cleavage means, one may consider the action of said enzymes on any of these structures, and on any other structures that fall within the definition of a cleavage structure. The cleavage sites indicated on the structures in Fig. 15 are presented by way of example. Specific cleavage at any site within such a structure is contemplated.

Improvements in an enzyme may be an increased or decreased rate of cleavage of one or more types of structures. Improvements may also result in more or fewer sites of cleavage on one or more of said cleavage structures. In developing a library of new structure-specific nucleases for use in nucleic acid cleavage assays, improvements may have many different embodiments, each related to the specific substrate structure used in a particular assay.

As an example, one embodiment of the Invader™-directed cleavage assay of the present invention may be considered. In the Invader™ directed cleavage assay, the accumulation of cleaved material is influenced by several features of the enzyme behavior. Not surprisingly, the turnover rate, or the number of structures that can be cleaved by a single enzyme molecule in a set amount of time, is very important in determining the amount of material processed during the course of an assay reaction. If an enzyme takes a long time to recognize a substrate (*e.g.*, if it is presented with a less-than-optimal structure), or if it takes a long time to execute cleavage, the rate of product accumulation is lower than if these steps proceeded quickly. If these steps are quick, yet the enzyme "holds on" to the cleaved structure, and does not immediately proceed to another uncut structure, the rate will be negatively affected.

Enzyme turnover is not the only way in which enzyme behavior can negatively affect the rate of accumulation of product. When the means used to visualize or measure product is specific for a precisely defined product, products that deviate from that definition may escape detection, and thus the rate of product accumulation may appear to be lower than it is. For example, if one had a sensitive detector for trinucleotides that could not see di- or tetranucleotides, or any sized oligonucleotide other that 3 residues, in the Invader™-directed cleavage assay of the present invention any errant cleavage would reduce the detectable signal proportionally. It can be seen from the cleavage data presented here that, while there is usually one site within a probe that is favored for cleavage, there are often products that arise from cleavage one or more nucleotides away from the primary cleavage site. These are products that are target dependent, and are thus not non-specific background. Nevertheless, if a subsequent visualization system can detect only the primary product, these represent a loss of signal. One example of such a selective visualization system is the charge reversal readout presented herein, in which the balance of positive and negative charges determines the behavior of the products. In such a system the presence of an extra nucleotide or the absence of an expected nucleotide can excluded a legitimate cleavage product from ultimate detection by leaving that product with the wrong balance of charge. It can be easily seen that any assay that can sensitively distinguish the nucleotide content of an oligonucleotide, such as standard stringent hybridization, suffers in sensitivity when some fraction of the legitimate product is not eligible for successful detection by that assay.

These discussions suggest two highly desirable traits in any enzyme to be used in the method of the present invention. First, the more rapidly the enzyme executes an entire cleavage reaction, including recognition, cleavage and release, the more signal it may potentially created in the Invader™-directed cleavage assay. Second, the more successful an enzyme is at focusing on a single cleavage site within a structure, the more of the cleavage product can be successfully detected in a selective read-out.

The rationale cited above for making improvements in enzymes to be used in the Invader™-directed cleavage assay are meant to serve as an example of one direction in which improvements might be sought, but not as a limit on either the nature or the applications of improved enzyme activities. As another direction of activity change that would be appropriately considered improvement, the DNAP-associated 5' nucleases may be used as an example. In creating some of the polymerase-deficient 5' nucleases described herein it was found that the those that were created by deletion of substantial portions of the polymerase

domain, as depicted in Fig. 4, assumed activities that were weak or absent in the parent proteins. These activities included the ability to cleave the non-forked structure shown in Fig. 15D, a greatly enhanced ability to exonucleolytically remove nucleotides from the 5' ends of duplexed strands, and a nascent ability to cleave circular molecules without benefit of a free 5' end.

In addition to the 5' nucleases derived from DNA polymerases, the present invention also contemplates the use of structure-specific nucleases that are not derived from DNA polymerases. For example, a class of eukaryotic and archaebacterial endonucleases have been identified which have a similar substrate specificity to 5' nucleases of Pol I-type DNA polymerases. These are the FEN1 (Flap EndoNuclease), RAD2, and XPG (Xeroderma Pigmentosa-complementation group G) proteins. These proteins are involved in DNA repair, and have been shown to favor the cleavage of structures that resemble a 5' arm that has been displaced by an extending primer during polymerization, similar to the model depicted in Fig. 15B. Similar DNA repair enzymes have been isolated from single cell and higher eukaryotes and from archaea, and there are related DNA repair proteins in eubacteria. Similar 5' nucleases have also be associated with bacteriophage such as T5 and T7.

Recently, the 3-dimensional structures of DNAPTaq and T5 phage 5'-exonuclease (Fig. 58) were determined by X-ray diffraction (Kim *et al.*, Nature 376:612 [1995]; and Ceska *et al.*, Nature 382:90 [1995]). The two enzymes have very similar 3-dimensional structures despite limited amino acid sequence similarity. The most striking feature of the T5 5'-exonuclease structure is the existence of a triangular hole formed by the active site of the protein and two alpha helices (Fig. 58). This same region of DNAPTaq is disordered in the crystal structure, indicating that this region is flexible, and thus is not shown in the published 3-dimensional structure. However, the 5' nuclease domain of DNAPTaq is likely to have the same structure, based its overall 3-dimensional similarity to T5 5'-exonuclease, and that the amino acids in the disordered region of the DNAPTaq protein are those associated with alpha helix formation. The existence of such a hole or groove in the 5' nuclease domain of DNAPTaq was predicted based on its substrate specificity (Lyamichev *et al.*, *supra*).

It has been suggested that the 5' arm of a cleavage structure must thread through the helical arch described above to position said structure correctly for cleavage (Ceska *et al.*, *supra*). One of the modifications of 5' nucleases described herein opened up the helical arch portion of the protein to allow improved cleavage of structures that cut poorly or not at all (*e.g.*, structures on circular DNA targets that would preclude such threading of a 5' arm).

The gene construct that was chosen as a model to test this approach was the one called Cleavase® BN, which was derived from DNAPTaq but does not contain the polymerase domainn (Ex. 2). It comprises the entire 5' nuclease domain of DNAP Taq, and thus should be very close in structure to the T5 5' exonuclease. This 5' nuclease was chosen to demonstrate the principle of such a physical modification on proteins of this type. The arch-opening modification of the present invention is not intended to be limited to the 5' nuclease domains of DNA polymerases, and is contemplated for use on any structure-specific nuclease which includes such an aperture as a limitation on cleavage activity. The present invention contemplates the insertion of a thrombin cleavage site into the helical arch of DNAPs derived from the genus *Thermus* as well as 5' nucleases derived from DNAPs derived from the genus *Thermus*. The specific example shown herein using the Cleavase® BN/thrombin nuclease merely illustrates the concept of opening the helical arch located within a nuclease domain. As the amino acid sequence of DNAPs derived from the genus *Thermus* are highly conserved, the teachings of the present invention enable the insertion of a thrombin site into the helical arch present in these DNAPs and 5' nucleases derived from these DNAPs.

The opening of the helical arch was accomplished by insertion of a protease site in the arch. This allowed post-translational digestion of the expressed protein with the appropriate protease to open the arch at its apex. Proteases of this type recognize short stretches of specific amino acid sequence. Such proteases include thrombin and factor Xa. Cleavage of a protein with such a protease depends on both the presence of that site in the amino acid sequence of the protein and the accessibility of that site on the folded intact protein. Even with a crystal structure it can be difficult to predict the susceptibility of any particular region of a protein to protease cleavage. Absent a crystal structure it must be determined empirically.

In selecting a protease for a site-specific cleavage of a protein that has been modified to contain a protease cleavage site, a first step is to test the unmodified protein for cleavage at alternative sites. For example, DNAPTaq and Cleavase® BN nuclease were both incubated under protease cleavage conditions with factor Xa and thrombin proteases. Both nuclease proteins were cut with factor Xa within the 5' nuclease domain, but neither nuclease was digested with large amounts of thrombin. Thus, thrombin was chosen for initial tests on opening the arch of the Cleavase® BN enzyme.

In the protease/Cleavase® modifications described herein the factor Xa protease cleaved strongly in an unacceptable position in the unmodified nuclease protein, in a region

likely to compromise the activity of the end product. Other unmodified nucleases contemplated herein may not be sensitive to the factor Xa, but may be sensitive to thrombin or other such proteases. Alternatively, they may be sensitive to these or other such proteases at sites that are immaterial to the function of the nuclease sought to be modified. In approaching any protein for modification by addition of a protease cleavage site, the unmodified protein should be tested with the proteases under consideration to determine which proteases give acceptable levels of cleavage in other regions.

Working with the cloned segment of DNAPTaq from which the Cleavase® BN protein is expressed, nucleotides encoding a thrombin cleavage site were introduced in-frame near the sequence encoding amino acid 90 of the nuclease gene. This position was determined to be at or near the apex of the helical arch by reference to both the 3-dimensional structure of DNAPTaq, and the structure of T5 5' exonuclease. The encoded amino acid sequence, LVPRGS, was inserted into the apex of the helical arch by site-directed mutagenesis of the nuclease gene. The proline (P) in the thrombin cleavage site was positioned to replace a proline normally in this position in Cleavase® BN because proline is an alpha helix-breaking amino acid, and may be important for the 3-dimensional structure of this arch. This construct was expressed, purified and then digested with thrombin. The digested enzyme was tested for its ability to cleave a target nucleic acid, bacteriophage M13 genomic DNA, that does not provide free 5' ends to facilitate cleavage by the threading model.

While the helical arch in this nuclease was opened by protease cleavage, it is contemplated that a number of other techniques could be used to achieve the same end. For example, the nucleotide sequence could be rearranged such that, upon expression, the resulting protein would be configured so that the top of the helical arch (amino acid 90) would be at the amino terminus of the protein, the natural carboxyl and amino termini of the protein sequence would be joined, and the new carboxyl terminus would lie at natural amino acid 89. This approach has the benefit that no foreign sequences are introduced and the enzyme is a single amino acid chain, and thus may be more stable that the cleaved 5' nuclease. In the crystal structure of DNAPTaq, the amino and carboxyl termini of the 5'-exonuclease domain lie in close proximity to each other, which suggests that the ends may be directly joined without the use of a flexible linker peptide sequence as is sometimes necessary. Such a rearrangement of the gene, with subsequent cloning and expression could be accomplished by standard PCR recombination and cloning techniques known to those skilled in the art.

The present invention also contemplates the use of nucleases isolated from a organisms that grow under a variety of conditions. The genes for the FEN-1/XPG class of enzymes are found in organisms ranging from bacteriophage to humans to the extreme thermophiles of Kingdom Archaea. For assays in which high temperature is to be used, it is contemplated that enzymes isolated from extreme thermophiles may exhibit the thermostability required of such an assay. For assays in which it might be desirable to have peak enzyme activity at moderate temperature or in which it might be desirable to destroy the enzyme with elevated temperature, those enzymes from organisms that favor moderate temperatures for growth may be of particular value.

An alignment of a collection of FEN-1 proteins sequenced by others is shown in Figs. 59A-E (SEQ ID NOS:135-145). It can be seen from this alignment that there are some regions of conservation in this class of proteins, suggesting that they are related in function, and possibly in structure. Regions of similarity at the amino acid sequence level can be used to design primers for *in vitro* amplification (PCR) by a process of back translating the amino acid sequence to the possible nucleic acid sequences, then choosing primers with the fewest possible variations within the sequences. These can be used in low stringency PCR to search for related DNA sequences. This approach permits the amplification of DNA encoding a FEN-1 nuclease without advance knowledge of the actual DNA sequence.

It can also be seen from this alignment that there are regions in the sequences that are not completely conserved. The degree of difference observed suggests that the proteins may have subtle or distinct differences is substrate specificity. In other words, they may have different levels of cleavage activity on the cleavage structures of the present invention. When a particular structure is cleaved at a higher rate than the others, this is referred to a preferred substrate, while a structure that is cleaved slowly is considered a less preferred substrate. The designation of preferred or less preferred substrates in this context is not intended to be a limitation of the present invention. It is contemplated that some embodiments the present invention will make use of the interactions of an enzyme with a less preferred substrate. Candidate enzymes are tested for suitability in the cleavage assays of the present invention using the assays described below.

1. Structure Specific Nuclease Assay

Testing candidate nucleases for structure-specific activities in these assays is done in much the same way as described for testing modified DNA polymerases in Example 2, but

with the use of a different library of model structures. In addition to assessing the enzyme performance in primer-independent and primer-directed cleavage, a set of synthetic hairpins are used to examine the length of duplex downstream of the cleavage site preferred by the enzyme.

The FEN-1 and XPG 5' nucleases used in the present invention must be tested for activity in the assays in which they are intended to be used, including but not limited to the Invader™-directed cleavage detection assay of the present invention and the CFLP® method of characterizing nucleic acids (the CFLP® method is described in co-pending Application Serial Nos. 08/337,164, 08/402,601, 08/484,956 and 08/520,946; the disclosures of these applications are incorporated herein by reference). The Invader™ assay uses a mode of cleavage that has been termed "primer directed" of "primer dependent" to reflect the influence of the an oligonucleotide hybridized to the target nucleic acid upstream of the cleavage site. In contrast, the CFLP® reaction is based on the cleavage of folded structure, or hairpins, within the target nucleic acid, in the absence of any hybridized oligonucleotide. The tests described herein are not intended to be limited to the analysis of nucleases with any particular site of cleavage or mode of recognition of substrate structures. It is contemplated that enzymes may be described as 3' nucleases, utilizing the 3' end as a reference point to recognize structures, or may have a yet a different mode of recognition. Further, the use of the term 5' nucleases is not intended to limit consideration to enzymes that cleave the cleavage structures at any particular site. It refers to a general class of enzymes that require some reference or access to a 5' end to effect cleavage of a structure.

A set of model cleavage structures have been created to allow the cleavage ability of unknown enzymes on such structures to be assessed. Each of the model structures is constructed of one or more synthetic oligonucleotides made by standard DNA synthesis chemistry. Examples of such synthetic model substrate structures are shown in Figs. 26 and 60. These are intended only to represent the general folded configuration desirable is such test structures. While a sequence that would assume such a structure is indicated in the Figures, there are numerous other sequence arrangements of nucleotides that would be expected to fold in such ways. The essential features to be designed into a set of oligonucleotides to perform the tests described herein are the presence or absence of a sufficiently long 3' arm to allow hybridization of an additional nucleic acid to test cleavage in a "primer-directed" mode, and the length of the duplex region. In the set depicted in Fig. 60, the duplex lengths of the S-33 and the 11-8-0 structures are 12 and 8 basepairs, respectively.

This difference in length in the test molecules facilitates detection of discrimination by the candidate nuclease between longer and shorter duplexes. Additions to this series expanding the range of duplex molecules presented to the enzymes, both shorter and longer, may be used. The use of a stabilizing DNA tetraloop (Antao *et al.*, Nucl. Acids Res., 19:5901 [1991]) or triloop (Hiraro *et al.*, Nuc. Acids Res., 22:576 [1994]) at the closed end of the duplex helps ensure formation of the expected structure by the oligonucleotide.

The model substrate for testing primer directed cleavage, the "S-60 hairpin" (SEQ ID NO:40) is described in Example 11. In the absence of a primer this hairpin is usually cleaved to release 5' arm fragments of 18 and 19 nucleotides length. An oligonucleotide, termed P-14 (5'-CGAGAGACCACGCT-3'; SEQ ID NO:108), that extends to the base of the duplex when hybridized to the 3' arm of the S-60 hairpin gives cleavage products of the same size, but at a higher rate of cleavage.

To test invasive cleavage a different primer is used, termed P-15 (5'-CGAGAGACCACGCTG-3'; SEQ ID NO:30). In a successful invasive cleavage the presence of this primer shifts the site of cleavage of S-60 into the duplex region, usually releasing products of 21 and 22 nucleotides length.

The S-60 hairpin may also be used to test the effects of modifications of the cleavage structure on either primer-directed or invasive cleavage. Such modifications include, but are not limited to, use of mismatches or base analogs in the hairpin duplex at one, a few or all positions, similar disruptions or modifications in the duplex between the primer and the 3' arm of the S-60, chemical or other modifications to one or both ends of the primer sequence, or attachment of moieties to, or other modifications of the 5' arm of the structure. In all of the analyses using the S-60 or a similar hairpin described herein, activity with and without a primer may be compared using the same hairpin structure.

The assembly of these test reactions, including appropriate amounts of hairpin, primer and candidate nuclease are described in Example 2. As cited therein, the presence of cleavage products is indicated by the presence of molecules which migrate at a lower molecular weight than does the uncleaved test structure. When the reversal of charge of a label is used the products will carry a different net charge than the uncleaved material. Any of these cleavage products indicate that the candidate nuclease has the desired structure-specific nuclease activity. By "desired structure-specific nuclease activity" it is meant only that the candidate nuclease cleaves one or more test molecules. It is not

necessary that the candidate nuclease cleave at any particular rate or site of cleavage to be considered successful cleavage.

VIII. Signal Enhancement By Completion Of An Activated Protein Binding Site

In addition to the DNA polymerase tailing reaction described above, the present invention also contemplates the use of the products of the invasive cleavage reaction to form activated protein binding sites, such as RNA polymerase promoter duplexes, thereby allowing the interaction of the completed site to be used as an indicator of the presence of the nucleic acid that is the target of the invasive cleavage reaction. By way of example, when an RNA polymerase promoter duplex is activated by being made complete (*i.e.*, double-stranded over that portion of the promoter region required for polymerase binding) through the hybridization of the oligonucleotide product of the invasive cleavage reaction, the synthesis of RNA can be used as such an indicator.

It is not intended that the transcription reaction of the present invention be limited to the use of any particular RNA polymerase or RNA polymerase promoter region. Promoter sequences are well characterized for several bacteriophage, including bacteriophage SP6, T7 and T3. In addition, promoter sequences have been well characterized for a number of both eukaryotic and prokaryotic RNA polymerases. In a preferred embodiment, the promoter used enables transcription from one of the bacteriophage RNA polymerases. In a particularly preferred embodiment, the promoter used enables transcription by T7 RNA polymerase. Means of performing transcription *in vitro* are well known in the art and commercial kits are available for performing transcription with eukaryotic, prokaryotic or bacteriophage RNA polymerases (*e.g.*, from Promega).

The protein binding regions of the present invention are not limited to the bacteriophage RNA polymerase promoters described above. Other promoter sequences that are contemplated are those of prokaryotes and eukaryotes. For example, many strains of bacteria and fungi are used for the expression of heterologous proteins. The minimal promoters required for transcription by the RNA polymerases of organisms such as yeast and other fungi, eubacteria, nematodes, and cultured mammalian cells are well described in the literature and in the catalogs of commercial suppliers of DNA vectors for the expression of foreign proteins in these organisms.

The binding sites for other types of nucleic acid (*e.g.*, DNA) binding proteins are contemplated for use in the present invention. For example, proteins involved in the

regulation of genes exert their effects by binding to the DNA in the vicinity of the promoter from which the RNA from that gene is transcribed. The *lac* operator of *E. coli* is one example of a particularly well characterized and commonly used gene regulation system in which the *lac* repressor protein binds to specific sequences that overlap, and thus block, the promoter for the genes under the repressor's control (Jacob and Monod, Cold Spring Harbor Symposium on Quantitative Biol. XXVI:193-211 [1961]). Many similar systems have been described in bacteria, including the *trp* and *AraC* regulatory systems. Given the large amount of information available about bacterial promoters, the steps described below for the design of suitable partial promoters for the bacteriophage RNA polymerases can be readily adapted to the design of detection systems based on these other promoters.

As noted above, many of the bacterial promoters are under the control of a repressor or other regulatory protein. It is considered to be within the scope of the present invention to include the creation of composite binding sites for these regulatory proteins through the provision of a nucleic acid fragment (*e.g.*, a non-target cleavage product generated in an invasive cleavage reaction). The binding of the regulatory protein to the completed protein binding region (*e.g.*, the composite binding region) can be assessed by any one of a number of means, including slowed electrophoretic migration of either the protein or the DNA fragment, or by a conformational change in the protein or DNA upon binding. In addition, transcription from a downstream promoter can be monitored for up- or down-regulation as a result of the binding of the regulatory protein to the completed protein binding region.

In addition to the bacterial systems described above, many genes in eukaryotic systems have also been found to be under the control of specific proteins that bind to specific regions of duplex DNA. Examples include, but are not limited to, the OCT-1, OCT-2 and AP-4 proteins in mammals and the GAL4 and GCN4 proteins in yeast. Such regulatory proteins usually have a structural motif associated with duplex nucleic acid binding, such as a helix-turn-helix, a zinc finger or a leucine zipper [for review, *see*, *Molecular and Cellular Biology*, Wolfe (Ed.), Wadsworth Publishing Co., Belmont, CA, pp. 694-715 [1993]).

For simplicity the test reaction described here will refer to T7 RNA polymerase, and its promoter. This is not intended to limit the invention to the use of this RNA polymerase, and those skilled in the art of molecular biology would be able to readily adapt this described test to the examination of any of the DNA binding proteins, RNA polymerases and their binding or promoter sites discussed above.

It is known in the art that active T7 promoters can be formed by the hybridization of two oligonucleotides, each comprising either the top or bottom strand of the promoter sequence, such that a complete un-nicked duplex promoter is formed (Milligan *et al.*, Nucl. Acids Res., 15:21, 8783-8798 (1987)]. The present invention shows that one way of making the initiation of transcription dependent on the products of an invasive cleavage reaction is to design the probe for the cleavage reaction such that a portion of an RNA polymerase promoter is released as product. The remaining DNA piece or pieces required to assemble a promoter duplex may either be provided as elements in the reaction mixture, or they may be produced by other invasive cleavage events. If the oligonucleotide pieces are designed to comprise appropriate regions of complementarity they may base pair to form a complete promoter duplex composed of three or more nucleic acid fragments, as depicted in Fig. 88B. A promoter assembled in this way will have nicks in the backbone of one or both strands. In one embodiment, these nicks may be covalently closed through the use of a DNA ligase enzyme. In a preferred embodiment, the nicks are positioned such that transcription can proceed without ligation. In selecting the site of a nick created by the assembly of the partial promoter fragment, at least one nick should be within the recognized promoter region for the RNA polymerase to be used. When a bacteriophage promoter is used, a nick should be between nucleotides -17 and -1, measured from the site of transcription initiation at +1. In a preferred embodiment, a nick will be between nucleotides -13 and -8. In a particularly preferred embodiment, a nick will be between nucleotides -12 and -10 on the non-template strand of the bacteriophage promoter.

When nicks are to be left unrepaired (*i.e.*, not covalently closed with a DNA ligase) it is important to assess the effect of the nick location on the level of transcription from the assembled promoter. A simple test is to combine the oligonucleotides that comprise the separate portions of the promoter with an oligonucleotide that comprises one entire strand of the promoter to be assembled, thereby forming a duplex promoter with a nick in one strand. If the nick is in the top, or non-template strand of the promoter, then the oligonucleotide that comprises the complete promoter is made to include additional non-promoter sequence on its 5' end to serve as a template to be copied in the transcription. This arrangement is depicted in Fig. 88B. Alternatively, if the nick is to be in the bottom, or template strand of the promoter, then the partial promoter oligonucleotide that covers the +1 position, the transcription start site, will include the additional template sequence. This arrangement is depicted in Figs. 95A-D (this Figure shows several different embodiments in which a cut

probe or non-target cleavage product is used to form a composite promoter which contains one or more nicks on the template strand). In either case, the separate oligonucleotides are combined to form the complete promoter, and the assembly is used in a transcription reaction to create RNA.

To measure the effect of the nick, a substantially identical promoter fragment is created by hybridization of two oligonucleotides that each comprise one strand of the full-length promoter to create an un-nicked version of the same promoter. These two molecular assemblies are tested in parallel transcription reactions and the amount of the expected RNA that is produced in each reaction is measured for both size and yield. A preferred method of assessing the size of the RNA is by electrophoresis with subsequent visualization. If a labeled nucleotide (*e.g.*, ^{32}P- GTP, or fluorescein-UTP) is used in the transcription, the RNA can be detected and quantitated by autoradiography, fluorescence imaging or by transfer to support membrane with subsequent detection (*e.g.*, by antibody or hybridization probing). Alternatively, if unlabeled RNA is produced the amounts may be determined by other methods known in the art, such as by spectrophotometry or by electrophoresis with subsequent staining and comparison to known standards.

If the size of the RNA is as predicted by the template sequence, or if it matches that produced from the control promoter, it can be presumed to have initiated transcription at the same site in the complex, and to have produced essentially the same RNA product. If the product is much shorter then transcription is either initiating at an internal site or is terminating prematurely (Schenborn and Mierendorf, Nucl. Acids Res., 13:17, 6223 [1985]; and Milligan *et al.*, *supra.*). While this does not indicate that the assembly tested is completely unsuitable for the assay, the partial transcripts will reduce the gross amount of RNA created, perhaps compromising the signal from the assay, and such products would require further characterization (*e.g.*, finger printing or sequencing) to identify the nucleotide content of the product. It is preferred that the size of the RNA produced matches that of the RNA produced in the control reaction.

The yield of the reaction is also examined. It is not necessary that the level of transcription matches that of the control reaction. In some instances (see Ex. 41, below) the nicked promoter may have an enhanced rate of transcription, while in other arrangements transcription may be reduced (relative to the rate from the un-nicked promoter assembly). It is only required that the amount of product be within the detection limits of the method to be used with the test promoter.

It is reported that transcription from a bacteriophage promoter can produce 200 to 1000 copies of each transcription template (template plus active promoter) in a reaction. These levels of transcription are not required by the present invention. Reactions in which one RNA is produced for each template are also contemplated.

The test described above will allow a promoter with a nick in any position to be assessed for utility in this assay. It is an objective of this invention to provide one or more of the oligos which comprise a partial promoter region through invasive cleavage event(s). In this embodiment, the partial promoter sequences are attached to the probe oligonucleotide in the invasive cleavage assay, and are released by cleavage at specific site, as directed by the Invader™ oligonucleotide. It is also intended that transcription be very poor or nonexistent in the absence of the correctly cleaved probe. To assess the success of any oligonucleotide design at meeting these objectives, several transcription reaction tests can be performed.

For a promoter assembly that will have a nick on the non-template strand, several partial assemblies that should be tested are shown in Figs. 86 A-D. By way of example, but not by way of limitation, this Figure depicts the tests for a nicked promoter in which the upstream, or 5' portion of the non-template strand is to be provided by the invasive cleavage assay. This fragment is seen in Fig. 86A labeled as "cut probe". Transcription reactions incubated in the presence of the duplex shown in Fig. 86A will test the ability of the upstream partial promoter to allow initiation of transcription when hybridized to a bottom strand, termed a "copy template." Similarly, a reaction performed in the presence of the duplex depicted in Fig. 86B will test the ability of the partial promoter fragment nearest the initiation site (the +1 site, as indicated in Fig. 85B) to support transcription of the copy template. It is an important feature of the present invention that neither of these partial promoter duplexes be able to support transcription at the same level as would by seen in transcription from an intact promoter as depicted in Fig. 85B. It is preferred that neither of these partial promoters be sufficient to initiate detectable transcription in the time course of an average transcription reaction (*i.e.*, within about an hour of incubation).

Figs. 86C and 86D depict two other duplex arrangements designed to test the effect of uncut probe within the transcription reaction. Fig. 86C depicts the duplex formed between only the uncut probe and the copy template, while Fig. 86D includes the other portion of the promoter. The 3' region of the probe is not complementary to the promoter sequence and therefore produces an unpaired branch in the middle of the promoter. It is an important feature of the present invention that neither of these branched promoter duplexes be able to

support transcription at the same level as would by seen in transcription from an intact promoter as depicted in Fig. 85B. It is preferred that neither of these branched promoters be sufficient to initiate detectable transcription in the time course of an average transcription reaction (*i.e.*, within about an hour of incubation).

In one embodiment of the transcription system of the present invention, the initiation of transcription from the copy template in the absence of a complete promoter, or in the presence of a branched promoter, is prevented by the judicious placement of the nick or nicks in the composite promoter. For example, as shown in the examples below, placement of a nick between the -12 and -11 nucleotides of the non-template strand of the bacteriophage T7 promoter allows transcription to take place only when the probe has been successfully cut, as in an invasive cleavage reaction. However, in some instances where the invasive cleavage reaction is to provide the upstream portion of the non-template strand of the promoter (*e.g.*, as depicted in Fig. 88B) it may be necessary or desirable to place the nick on that strand in a particular position for reasons other than providing an optimal composite promoter (*i.e.*, one that is inactive in the absence of any one of the promoter pieces). It may be necessary or desirable to place the nick in such a way that the creation of a branched complete promoter (Fig. 86D) has an undesirable level of transcription, reducing dependence of RNA production on the success of the invasive cleavage step. It is shown in the examples below that transcription from such a branched promoter can be suppressed by a modification of the downstream non-template promoter piece, shown as the "Partial Promoter Oligonucleotide" in Figs. 86, 88, 90 and 95D. As depicted in Fig. 90, the partial promoter oligonucleotide can be provided with a 5' "tail" of nucleotides that are not complementary to the template strand of the promoter, but which are complementary to the 3' portion of the probe oligonucleotide that would be removed in the invasive cleavage reaction. When uncut probe hybridizes to the copy template with the bound 5' tailed partial promoter oligonucleotide, the 5' tail can basepair to the 3' region of the probe, forming a three-way junction as depicted in Fig. 90A. This can effectively shut off transcription, as shown below. When a cut probe hybridizes, as shown in Fig. 90B, a promoter with a small branch is formed, and it is shown herein that such a branched promoter can initiate transcription. Furthermore, if care is taken in selecting the sequence of the 5' tail (*i.e.*, if the first unpaired base is the same nucleotide at the 3' nucleotide of the cut probe, so that they compete for hybridization to the same template strand base), the resulting branched structure may also be cleaved by one of the structure

specific nucleases of the present invention, creating the un-branched promoter depicted in Fig. 90C, in some instances enhancing transcription over that seen with the Fig. 90B promoter.

The promoter duplex that is intended to be created, in this embodiment, by the successful execution of the Invader™ directed cleavage assay will include both the "cut probe" and the partial promoter oligonucleotide depicted in Figs. 86A and B, aligned on a single copy template nucleic acid. The testing of the efficiency of transcription of such a nicked promoter segment in comparison to the intact promoter is described above. All of the oligonucleotides described for these test molecules may be created using standard synthesis chemistries.

The set of test molecules depicted in Fig. 86 is designed to assess the transcription capabilities of the variety of structures that may be present in reactions in which the 5' portion of the non-template strand of the promoter is to be supplied by the Invader™ directed cleavage. It is also envisioned that a different portion of partial promoter may be supplied by the invasive cleavage reaction (*e.g.*, the downstream segment of the non-template strand of the promoter), as is shown in Fig. 94. Portions of the template strand of the promoter may also be provided by the cut probe, as shown in Figs. 95A-D. An analogous set of test molecules, including "cut" and uncut versions of the probe to be used in the invasive cleavage assay may be created to test any alternative design, whether the nick is to be located on the template or non template strand of the promoter.

The transcription-based visualization methods of the present invention may also be used in a multiplex fashion. Reactions can be constructed such that the presence of one particular target leads to transcription from one type of promoter, while the presence of a different target sequence (*e.g.*, a mutant or variant) or another target suspected of being present, may lead to transcription from a different (*i.e.*, a second) type of promoter. In such an embodiment, the identity of the promoter from which transcription was initiated could be deduced from the type or size of the RNA produced.

By way of example, but not by way of limitation, the bacteriophage promoters can be compared with such an application in view. The promoters for the phage T7, T3 and SP6 are quite similar, each being about 15 to 20 basepairs long, and sharing about 45% identity between -17 and -1 nucleotides, relative to the start of transcription. Despite these similarities, the RNA polymerases from these phage are highly specific for their cognate promoters, such that the other promoters may be present in a reaction, but will not be transcribed (Chamberlin and Ryan, Enzymes XV:87-108 [1982]). Because these promoters

are similar in size and in the way in which they are recognized by their polymerases (Li *et al.*, Biochem. 35:3722 [1996]) similar nicked versions of the promoters may be designed for use in the methods of the present invention by analogy to the examples described herein which employ the T7 promoter. Because of the high degree of specificity of the RNA polymerases, these nicked promoters may be used together to detect multiple targets in a single reaction. There are many instances in which it would be highly desirable to detect multiple nucleic acid targets in a single sample, including cases in which multiple infectious agents may be present, or in which variants of a single type of target may need to be identified. Alternatively, it is often desirable to use a combination of probes to detect both a target sequence and an internal control sequence, to gauge the effects of sample contaminants on the output of the assay. The use of multiple promoters allows the reaction to be assessed for both the efficiency of the invasive cleavage and the robustness of the transcription.

As stated above, the phage promoters were described in detail as an example of suitable protein binding regions (*e.g.*, which can be used to generate a composite promoter) for use in the methods of the present invention. The invention is not limited to the use of phage RNA polymerase promoter regions, in particular, and RNA polymerase promoter regions, in general. Suitably specific, well characterized promoters are also found in both prokaryotic and eukaryotic systems.

The RNA that is produced in a manner that is dependent of the successful detection of the target nucleic acid in the invasive cleavage reaction may be detected in any of several ways. If a labeled nucleotide is incorporated into the RNA during transcription, the RNA may be detected directly after fractionation (*e.g.*, by electrophoresis or chromatography). The labeled RNA may also be captured onto a solid support, such as a microtitre plate, a bead or a dipstick (*e.g.*, by hybridization, antibody capture, or through an affinity interaction such as that between biotin and avidin). Capture may facilitate the measuring of incorporated label, or it may be an intermediate step before probe hybridization or similar detection means. If the maximum amount of label is desired to be incorporated into each transcript, it is preferred that the copy template be very long, around 3 to 10 kilobases, so that each RNA molecule will carry many labels. Alternatively, it may be desired that a single site or a limited number of sites within the transcript be specifically labeled. In this case, it may be desirable to have a short copy template with only one or a few residues that would allow incorporation of the labeled nucleotide.

The copy template may also be selected to produce RNAs that perform specified functions. In a simple case, if an duplex-dependent intercalating fluorophore is to be used to detect the RNA product, it may be desirable to transcribe an RNA that is known to form duplexed secondary structures, such as a ribosomal RNA or a tRNA. In another embodiment, the RNA may be designed to interact specifically, or with particular affinity, with a different substance. It has been shown that a process of alternating steps of selection (*e.g.*, by binding to a target substance) and *in vitro* amplification (*e.g.*, by PCR) can be used to identify nucleic acid ligands with novel and useful properties (Tuerk and Gold, Science 249:505 [1990]). This system has been used to identify RNAs, termed ligands or aptamers, that bind tightly and specifically to proteins and to other types of molecules, such as antibiotics (Wang *et al.*, Biochem. 35:12338 [1996]) and hormones. RNAs can even be selected to bind to other RNAs through non-Watson-Crick interactions (Schmidt *et al.*, Ann. N.Y. Acad. Sci. 782:526 [1996]). A ligand RNA may be used to either inactivate or enhance the activity of a molecule to which it binds. Any RNA segment identified through such a process may also be produced by the methods of the present invention, so that the observation of the activity of the RNA ligand may be used as a specific sign of the presence of the target material in the invasive cleavage reaction. The ligand binding to its specific partner may also be used as another way of capturing a readout signal to a solid support.

The product RNA might also be designed to have a catalytic function (*e.g.*, to act as a ribozyme), allowing cleavage another molecule to be indicative of the success of the primary invasive cleavage reaction (Uhlenbeck, Nature 328:596 [1987]). In yet another embodiment, the RNA may be made to encode a peptide sequence. When coupled to an *in vitro* translation system (*e.g.*, the S-30 system derived from *E. coli* [Lesley, Methods Mol. Biol., 37:265 (1985)], or a rabbit reticulocyte lysate system [Dasso and Jackson, Nucleic Acids Res. 17:3129 (1989)], available from Promega), the production of the appropriate protein may be detected. In a preferred embodiment, the proteins produced include those that allow either colorimetric or luminescent detection, such as beta-galactosidase (lac-Z) or luciferase, respectively.

The above discussion focused on the use of the present transcription visualization methods in the context of the Invader™-directed cleavage assay (*i.e.*, the non-target cleavage products produced in the Invader™ assay were used to complete and activate a protein binding region, such as a promoter region). However, the transcription visualization methods are not limited to this context. Any assay which produces an oligonucleotide product having

relatively discrete ends can be used in conjunction with the present transcription visualization methods. For example, the homogenous assay described in U.S. Patent No. 5,210,015, particularly when conducted under conditions where polymerization cannot occur, produces short oligonucleotide fragments as the result of cleavage of a probe. If this assay is conducted under conditions where polymerase occurs, the site of cleavage of the probe may be focused through the use of nucleotide analogs that have uncleavable linkages at particular positions within the probe. These short oligonucleotides can be employed in a manner analogous to the cut probe or non-target cleavage products produced in the invasive cleavage reactions of the present invention. Additional assays which generate suitable oligonucleotide products are known to the art. For example, the non-target cleavage products produced in assays such as the "Cycling Probe Reaction" (Duck *et al.*, BioTech., 9:142 [1990] and U.S. Patents Nos. 4,876,187 and 5,011,769, herein incoporated by reference), in which shorter oligonucleotides are released from longer oligonucleotides after hybridization to a target sequence would be suitable, as would short restriction fragments released in assays where a probe is designed to be cleaved when successfully hybridized to an appropriate restriction recognition sequence (U.S. Patent No. 4,683,194, herein incorporated by reference).

Assays which generate short oligonucleotides having "ragged" (*i.e.*, not discrete) 3' ends can also be employed with success in the transcription reactions of the present invention when the oligonucleotide provided by this non-transcription reaction are used to provide a portion of the promoter region located downstream of the other oligonucleotide(s) which are required to complete the promoter region (that is a 3' tail or unpaired extension can be tolerated when the oligo is being used as the "Cut Probe" is in Figs. 94 and 95A).

IX. Signal Enhancement By Incorporating The Products Of An Invasive Cleavage Reaction Into A Subsequent Invasive Cleavage Reaction

As noted above, the oligonucleotide product released by the invasive cleavage can be used subsequently in any reaction or read-out method that uses oligonucleotides in the size range of a cleavage product. In addition to the reactions involving primer extension and transcription, described above, another enzymatic reaction that makes use of oligonucleotides is the invasive cleavage reaction. The present invention provide means of using the oligonucleotide released in a primary invasive cleavage reaction as a component to complete a cleavage structure to enable a secondary invasive cleavage reaction. One possible configuration of a primary cleavage reaction supplying a component for a secondary cleavage

structure is diagrammed in Fig. 96. Is not intended that the sequential use of the invasive cleavage product be limited to a single additional step. It is contemplated that many distinct invasive cleavage reactions may be performed in sequence.

The polymerase chain reaction uses a DNA replication method to create copies of a targeted segment of nucleic acid at a logarithmic rate of accumulation. This is made possible by the fact that when the strands of DNA are separated, each individual strand contains sufficient information to allow assembly of a new complementary strand. When the new strands are synthesized the number of identical molecules has doubled. Within 20 iterations of this process, the original may be copied 1 million-fold, making very rare sequences easily detectable. The mathematical power of a doubling reaction has been incorporated into a number of amplification assays, several of which are cited in Table 1.

By performing multiple, sequential invasive cleavage reactions the method of the present invention captures an exponential mathematical advantage without producing additional copies of the target analyte. In a simple invasive cleavage reaction the yield, Y, is simply the turnover rate, K, multiplied by the time of the reaction, t (*i.e.*, $Y = (K)(t)$). If Y is used to represent the yield of a simple reaction, then the yield of a compound (*i.e.*, a multiple, sequential reaction), assuming that each of the individual invasive cleavage steps has the same turnover rate, can be simply represented as Y^n, where n is the number of invasive cleavage reactions that have been performed in the series. If the yields of each step differ the ultimate yield can be represented as the product of the multiplication of the yields of each individual reaction in the series. For example, if a primary invasive cleavage reaction can produce one thousand products in 30 minutes, and each of those products can in turn participate in 1000 additional reactions, there will be 1000^2 copies (1000 x 1000) of the ultimate product in a second reaction. If a third reaction is added to the series, then the theoretical yield will be 1000^3 (1000 x 1000 x 1000). In the methods of the present invention the exponent comes from the number of invasive cleavage reactions in the cascade. This can be contrasted to the amplification methods described above (*e.g.*, PCR) in which Y is limited to 2 by the number of strands in duplex DNA, and the exponent n is the number of cycles performed, so that many iterations are necessary to accumulate large amounts of product.

To distinguish the exponential amplifications described above from those of the present invention, the former can be consider reciprocating reactions because the products the reaction feed back into the same reaction (*e.g.*, event one leads to some number of events 2, and each event 2 leads back to some number of events 1). In contrast, the events of the

present invention are sequential (*e.g.*, event 1 leads to some number of events 2; each event 2 leads to some number of events 3, etc., and no event can contribute to an event earlier in the chain).

The sensitivity of the reciprocating methods is also one of the greatest weaknesses when these assays are used to determine if a target nucleic acid sequence is present or absent in a sample. Because the product of these reactions is detectable copy of the starting material, contamination of a new reaction with the products of an earlier reaction can lead to false positive results, (*i.e.*, the apparent detection of the target nucleic acid in samples that do not actually contain any of that target analyte). Furthermore, because the concentration of the product in each positive reaction is so high, amounts of DNA sufficient to create a strong false positive signal can be communicated to new reactions very easily either by contact with contaminated instruments or by aerosol. In contrast to the reciprocating methods, the most concentrated product of the sequential reaction (*i.e.*, the product released in the ultimate invasive cleavage event, is not capable of initiating a like reaction or cascade if carried over to a fresh test sample). This is a marked advantage over the exponential amplification methods described above because the reactions of the present invention may be performed without the costly containment arrangements (*e.g.*, either by specialized instruments or by separate laboratory space) required by any reciprocating reaction. While the products of a penultimate event may be inadvertently transferred to produce a background of the ultimate product in the absence of the a target analyte, the contamination would need to be of much greater volume to give an equivalent risk of a false positive result.

When the term sequential is used it is not intended to limit the invention to configurations in which that one invasive cleavage reaction or assay must be completed before the initiation of a subsequent reaction for invasive cleavage of a different probe. Rather, the term refers to the order of events as would occur if only single copies of each of the oligonucleotide species were used in an assay. The primary invasive cleavage reaction refers to that which occurs first, in response to the formation of the cleavage structure on the target nucleic acid. Subsequent reactions may be referred to as secondary, tertiary and so forth, and may involve artificial "target" strands that serve only to support assembly of a cleavage structure, and which are unrelated to the nucleic acid analyte of interest.. While the complete assay may, if desired, be configured with each step of invasive cleavage separated either in space (*e.g.*, in different reaction vessels) or in time (*e.g.*, using a shift in reaction conditions, such as temperature, enzyme identity or solution condition, to enable the later

cleavage events), it is also contemplated that all of the reaction components may be mixed so that secondary reactions may be initiated as soon as product from a primary cleavage becomes available. In such a format, primary, secondary and subsequent cleavage events involving different copies of the cleavage structures may take place simultaneously.

Several levels of this sort of linear amplification can be envisioned, in which each successive round of cleavage produces an oligonucleotide that can participate in the cleavage of a different probe in subsequent rounds. The primary reaction would be specific for the analyte of interest with secondary (and tertiary, etc.) reactions being used to generate signal while still being dependent on the primary reaction for initiation.

The released product may perform in several capacities in the subsequent reactions. One of the possible variations are shown in Fig. 96, in which the product of one invasive cleavage reaction becomes the Invader™ oligonucleotide to direct the specific cleavage of another probe in a second reaction. In Fig. 96, the first invasive cleavage structure is formed by the annealing of the Invader ™ oligonucleotide ("Invader") and the probe oligonucleotide ("Probe 1") to the first target nucleic acid ("Target 1"). The target nucleic acid is divided into three regions based upon which portions of the Invader ™ and probe oligonucleotides are capable of hybridizing to the target (as discussed above and as shown in Fig. 25). Region 1 (region Y in Fig. 25) of the target has complementarity to only the Invader ™ oligonucleotide; region 3 (region Z in Fig. 25) of the target has complementarity to only the probe; and region 2 (region X in Fig. 25) of the target has complementarity to both the Invader ™ and probe oligonucleotides. It is noted that the sequential invasive cleavage reaction diagrammed in Fig. 96 employs an Invader™ and a probe oligonucleotide; the sequential cleavage reaction is not limited to the use of such a first cleavage structure. The first cleavage structure in the sequential reaction may also employ an Invader™ oligonucleotide, a mini probe and a stacker oligonucleotide as discussed in Section V above.

In Fig. 96, cleavage of Probe 1 releases the "Cut Probe 1" (indicated by the hatched line in both the cleaved and uncleaved Probe 1 in Fig. 96). The released Probe 1 is then used as the Invader ™ oligonucleotide in second cleavage. The second cleavage structure is formed by the annealing of the Cut Probe 1, a second probe oligonucleotide ("Probe 2") and a second target nucleic acid ("Target 2"). Probe 2 may be labelled (indicated by the star in Fig. 96) and detection of cleavage of the second cleavage structure may be accomplished by detecting the labelled cut Probe 2; the label may a radioisotope (*e.g.*, ^{32}P, ^{35}S), a fluorophore (*e.g.*, fluorescein), a reactive group capable of detection by a secondary agent (*e.g.*,

biotin/streptavidin), a positively charged adduct which permits detection by selective charge reversal (as discussed in Section IV above), etc. Alternatively, the cut Probe 2 may used in a tailing reaction (as discussed in Section VI above) or may used to complete or activate a protein binding site (as discussed in Section VIII above).

Another possible configuration for performing a sequential invasive cleavage reaction is diagrammed in Fig. 97. In this case, probe oligonucleotides that are cleaved in the primary reaction can be designed to fold back on themselves (*i.e.*, they contain a region of self-complementarity) to create a molecule that can serve as both the target and Invader™ oligonucleotide (termed here an "IT" complex). The IT complex then enables cleavage of a different probe present in the secondary reaction. Inclusion of an excess of the secondary probe molecule ("Probe 2"), allows each IT molecule to serve as the platform for the generation of multiple copies of cleaved secondary probe. In Fig. 97, the regions of self-complementarity contained within the 5' portion of the Invader™ oligonucleotide is indicated by the hatched ovals; the arrow between these two ovals indicates that these two regions can self-pair (as shown in the "Cut Probe 1"). The target nucleic acid is divided into three regions based upon which portions of the Invader ™ and probe oligonucleotides are capable of hybridizing to the target (as discussed above and it is noted that the target may be divided into four regions if a stacker oligonucleotide is employed). The second cleavage structure is formed by the annealing of the second probe ("Probe 2") to the fragment of Probe 1 ("Cut Probe 1") that was released by cleavage of the first cleavage structure. The Cut Probe 1 forms a hairpin or stem/loop structure near its 3' terminus by virtue of the annealing of the regions of self-complementarity contained within Cut Probe 1 (this self-annealed Cut Probe 1 forms the IT complex). The IT complex (Cut Probe 1) is divided into three regions. Region 1 of the IT complex has complementarity to the 3' portion of Probe 2; region 2 has complementarity to both the 3' end of Cut Probe 1 and to the 5' portion of Probe 2 (analogous to the region of overlap "X" shown in Fig. 25); and region 3 contains the region of self-complementarity (*i.e.*, region 3 is complementary to the 3' portion of the Cut Probe 1). Note that with regard to the IT complex (*i.e.*, Cut Probe 1), region 1 is located upstream of region 2 and region 2 is located upstream of region 3.

The cleavage products of the secondary invasive cleavage reaction (*i.e.*, Cut Probe 2) can either be detected, or can in turn be designed to constitute yet another integrated Invader™-target complex to be used with a third probe molecule, again unrelated to the preceding targets.

The present invention is not limited to the configurations diagrammed in Figs. 96 and 97. It is envisioned that the oligonucleotide product of a primary cleavage reaction may fill the role of any of the oligonucleotides described herein (*e.g.*, it may serve as a target strand without an attached Invader™ oligonucleotide-like sequence, or it may serve as a stacker oligonucleotide, as described above), to enhance the turnover rate seen in the secondary reaction by stabilizing the probe hybridization through coaxial stacking.

In a preferred embodiment, each subsequent reaction is enabled by (*i.e.*, is dependent upon) the product of the previous cleavage, so that the presence of the ultimate product may serve as an indicator of the presence of the target analyte. However, cleavage in the second reaction need not be dependent upon the presence of the product of the primary cleavage reaction; the product of the primary cleavage reaction may merely measurably enhance the rate of the second cleavage reaction.

In summary, the Invader™ assay cascade (*i.e.*, sequential invasive cleavage reactions) of the present invention is a combination of two or more linear assays that allow the accumulation of the ultimate product at an exponential rate, but without significant risk of carryover contamination.

The sequential invasive cleavage amplification of the present invention can be used as an intermediate boost to any of the detection methods (*e.g.*, gel based analysis by either standard or by charge reversal), polymerase tailing, and incorporation into a protein binding region, described herein. When used is such combinations the increased production of a specific cleavage product in the invasive cleavage assay reduces the burdens of sensitivity and specificity on the read-out systems, thus facilitating their use.

In addition to enabling a variety of detection platforms, the cascade strategy is suitable for multiplex analysis of individual analytes (*i.e.*, individual target nucleic acids) in a single reaction. The multiplex format can be categorized into two types. In one case, it is desirable to know the identity (and amount) of each of the analytes that can be present in a clinical sample, or the identity of each of the analytes as well as an internal control. To identify the presence of multiple individual analytes in a single sample, several distinct secondary amplification systems may be included. Each probe cleaved in response to the presence of a particular target sequence (or internal control) can be designed to trigger a different cascade coupled to different detectable moieties, such as different sequences to be extended by DNA polymerase or different dyes in an FET format. The contribution of each specific target

sequence to final product can thereby be tallied, allowing quantitative detection of different genes or alleles in a sample containing a mixture of genes/alleles.

In the second configuration, it is desirable to determine if any of several analytes are present in a sample, but the exact identity of each does not need to be known. For example, in blood banking it is desirable to know if any one of a host of infectious agents is present in a sample of blood. Because the blood is discarded regardless of which agent is present, different signals on the probes would not be required in such an application of the present invention, and may actually be undesirable for reasons of confidentiality. In this case, the 5' arms (*i.e.*, the 5' portion which will be released upon cleavage) of the different analyte-specific probes would be identical and would therefore trigger the same secondary signal cascade. A similar configuration would permit multiple probes complementary to a single gene to be used to boost the signal from that gene or to ensure inclusivity when there are numerous alleles of a gene to be detected.

In the primary Invader™ reaction, there are two potential sources of background. The first is from Invader™-independent cleavage of probe annealed to the target, to itself, or to one of the other oligonucleotides present in the reaction. It can be seen by consideration of Figs. 96 and 97 that the probes of the primary cleavage reactions depicted are designed to have regions of complementarity to the other oligonucleotides involved in the subsequent reactions, and, as depicted in Fig. 97, to other regions of the same molecule. The use of an enzyme that cannot efficiently cleave a structure that lacks a primer (*e.g.*, that cannot cleave the structures diagrammed in Fig. 16A or 16 D) is preferred for this reason. As shown in Figs. 99 and 100, the enzyme *Pfu* FEN-1 gives no detectable cleavage in the absence of the upstream oligonucleotide or even in the presence of an upstream oligonucleotide that fails to invade the probe-target complex. This indicates that the *Pfu* FEN-1 endonuclease is a suitable enzyme for use in the methods of the present invention.

Other structure-specific nucleases may be suitable as a well. As discussed in the first example, some 5' nucleases can be used in conditions that significantly reduce this primer -independent cleavage. For example, it has been shown that when the 5' nuclease of DNAPTaq is used to cleave hairpins the primer-independent cleavage is markedly reduced by the inclusion of a monovalent salt in the reaction (Lyamichev, *et al.*, [1993], *supra*).

Test For Invader™ Oligonucleotide-Independent Cleavage

A simple test can be performed for any enzyme in combination with any reaction buffer to gauge the amount of Invader™ oligonucleotide-independent cleavage to be expected from that combination. A small hairpin-like test molecule that can be used with or without a primer hybridized to a 3' arm, the S-60 molecule, is depicted in Fig. 30. The S-60 and the oligonucleotide P15 are a convenient set of molecules for testing the suitability of an enzyme for application in the present invention and conditions for using these molecules are described in Example 11. Other similar hairpins may be used, of a cleavage structure may be assembled from separate oligonucleotides as diagrammed in Figs 99a-e. Reactions using these structures to examine the activity of the *Pfu* FEN-1 enzyme in the presence or absence of an upstream overlapping oligonucleotide are described in Example 45 and the results are displayed in Fig. 100. To test any particular combination of enzyme and cleavage conditions, similar reactions can be assembled. Outside of the variables of reaction conditions to be tested for any particular enzyme (*e.g.*, salt sensitivities, divalent cation requirements) the test reactions should accommodate any known limitations of the test enzyme. For example, the test reactions should be performed at a temperature that is within the operating temperature range of the candidate enzyme, if known.

It is not necessary that multiple lengths of overlap be demonstrated for each candidate enzyme, but the activity of the enzyme in the absence of an upstream oligonucleotide (as shown in Fig. 99a) and in the presence of an oligonucleotide that does not overlap (Fig. 99b) should be assessed. It is preferable that structures lacking an upstream oligonucleotide be cleaved at less than one half of the rate seen in the presence of an upstream overlapping oligonucleotide. It is more preferable that these structures be cleaved at less than about on tenth the rate of the invasive cleavage structure. It is most preferred that cleavage of these structures occur at less than one percent the rate of the invasive cleavage structure.

If the cleaved product is to serve as an upstream oligonucleotide in a subsequent cleavage reaction, as diagrammed in Fig. 96, the most rapid reaction will be achieved if the other components of the second cleavage structure (*i.e.*, Target 2 and Probe 2 in Fig. 96) are provided in excess so that cleavage may proceed immediately after the upstream oligonucleotide (*i.e*, Cut Probe 1 in Fig. 96) is made available. To provide an abundance of the second target strand (Target 2 in Fig. 96) one may use an isolated natural nucleic acid, such as bacteriophage M13 DNA, or one may use a synthetic oligonucleotide. If a synthetic oligonucleotide is chosen as the second target sequence, the sequence employed must be

examined for regions of self-complementarity (similar considerations apply to short isolated natural nucleic acids such as restriction enzyme fragments or PCR products; natural nucleic acid targets whose 3' end is located ≥ 100 nucleotides downstream of the probe binding site on the target strand are sufficiently long enough to obviate the design considerations discussed below). Specifically, it must be determined that the 3' end of the synthetic oligonucleotide may not hybridize to the target strand (*i.e.*, intra-strand hybridization) upstream of the probe, triggering unintended cleavage. Simple examination of the sequence of the synthetic oligonucleotide should reveal if the 3' end has sufficient complementarity to the region of the target upstream of the probe binding site to pose a problem (*i.e*, it would reveal whether the synthetic oligonucleotide can form a hairpin at its 3' end which could act as an invading oligonucleotide to cause cleavage of the probe in the absence of the hybridization of the intended Invader™ oligonucleotide (*i.e.*, the cleavage product from the first invasive cleavage reaction). If 3 or more of the last 4 to 7 nucleotides (the 3' terminal region) of the synthetic target can basepair upstream of the probe such that there is an invasion into the probe-target duplex, or such that the duplexes formed by the synthetic target strand with its own 3' terminal region and with the probe abut without a gap and the 3' terminal region has an additional 1 or 2 nucleotides unpaired at the extreme 3' end of the synthetic target, then the sequence of the synthetic target oligonucleotide should be modified. The sequence may be changed to disrupt the interaction of the 3' terminal region or to increase the distance between the probe binding site and the regions to which the 3' terminus is binding. Alternatively, the 3' end may be modified to reduce its ability to direct cleavage (*e.g.*, by adding a 3' phosphate during synthesis) (see Ex. 35, Table 3) or by adding several additional nucleotides that will not basepair in a self-complementary manner (*i.e.*, they will not participate in the formation of a hairpin structure).

When the product of a first invasive cleavage reaction is designed to form a target which can fold on itself to direct cleavage of a second probe, the IT complex as diagrammed in Fig. 97, the design of the sequence used to form the stem/loop of the IT complex must be considered. To be factored into the design of such a probe are 1) the length of the region of self-complementarity, 2) the length of the region of overlap (region "X " in Fig. 25) and 3) the stability of the hairpin or stem/loop structure as predicted by both Watson-Crick base pairing and by the presence or absence of a particularly stable loop sequence (*e.g.*, a tetraloop [Tinoco *et al., supra*], or a triloop [Hirao *et al., supra*]). It is desirable that this sequence have nucleotides that can base pair (intrastrand), so that the second round of invasive cleavage

may occur, but that the structure not be so strong that its presence will prevent the cleavage of the probe in the primary reaction (*i.e.*, Probe 1 in Fig. 96). As shown herein, the presence of a secondary structure in the 5' arm of a cleavage structure cleaved by a structure-specific nuclease may inhibit cleavage by some structure-specific nucleases (Ex. 1).

The length of the region of self-complementarity within Probe 1 determines the length of the region of the duplex upstream of Probe 2 in the second cleavage structure (see Fig. 97). Different enzymes have different length requirements for this duplex to effect invasive cleavage efficiently. For example, the *Pfu* FEN-1 and *Mja* FEN-1 enzymes have been tested for the effect of this duplex length using the set of target/Invader™ oligonucleotide molecules depicted in Fig. 98 (*i.e.*, SEQ ID NOS:118, 119, 147-151). The invasive cleavage reactions were performed as described in Example 38, using 1 pM IT3 (SEQ ID NO:118), 2 μM probe PR1 (SEQ ID NO:119) for 5 min, and the rates of cleavage are shown in Table 2.

TABLE 2

Length of Duplex	*Pfu* FEN-1 Turnover, per min.	*Mja* FEN-1 Turnover, per min.
0	0	0
3	1	29
4	10	57
6	44	51
8	45	46

The data shown in Table 2 demonstrate that the *Pfu* FEN-1 enzyme can be used with stems of 3 or 4 bases, but that the rate of cleavage is maximized when the stem is greater than 4 basepairs in length. Table 2 shows that the *Mja* FEN-1 enzyme can cleave efficiently using shorter stems; however, as this enzyme can also cleave a probe in the absence of an upstream oligonucleotide, *Mja* FEN-1 is not preferred for use in the methods of the present invention.

A similar test can be performed using any candidate enzyme to determine how much self-complementarity may be designed into the Probe 1. The use of a shorter stem means that the overall probe may be shorter. This is beneficial because shorter probes are less costly to synthesize, and because shorter probes will have fewer sequences that might form unintended intrastrand structures. In assessing the activity of a candidate enzyme on the structures such as those shown in Fig. 98 it is not required that the stem length chosen allow the maximum rate of cleavage to occur. For example, in considering the case of *Pfu* FEN-1, the advantages

of using a 4 basepair stem (*e.g.*, cost or sequence limitations), with a cleavage rate of 10 cleavages per minute, may outweigh the rate advantage of using a longer 6 basepair stem (44 cleavages/min.), in the context of a particular experiment. It is within the scope of the present invention that some elements chosen for use in the assay be sub-optimal for performance of that particular element, if the use of a sub-optimal design benefits the objectives of that particular experiment as a whole.

In designing oligonucleotides to be employed as a probe that once cleaved forms a stem-loop structure as diagrammed in Fig. 97 (*i.e.*, Probe 1 in Fig. 97), it has been found that the stability of the loop is not a factor in the efficiency of cleavage of either Probe 1 or Probe 2. Loops tested have included stable triloops, loops of 3 and 4 nucleotides that were not predicted to be particularly stable (*i.e.*, the stability is determined by the duplex sequence and not by additional stabilizing interactions within the loop), and large loops of up to about 25 nucleotides.

X. Detection of Human Cytomegalovirus Viral DNA By Invasive Cleavage

Human cytomegalovirus (HCMV) causes, or is associated with, a wide variety of diseases in humans (Table 3). More than 90% of bone marrow or kidney transplant recipients (immunocompromised hosts) develop HCMV infections, most of which are due to reactivation of latent virus by immunosuppressive drugs, as well as transmission of virus by latently infected donor tissue or blood (Ackerman *et al.*, Transplant. Proc., 20(S1):468 [1988]; and Peterson *et al.*, Medicine 59:283 [1980]).

TABLE 3

Diseases Caused By Human Cytomegalovirus

cytomegalic inclusion
heterophil-negative disease in neonates
mononucleosis
interstitial pneumonia
pneumonitis
retinitis
hepatitis
pancreatitis
meningoencephalitis
gastrointestinal disease
disseminated infection

There are instances in which rapid, sensitive, and specific diagnosis of HCMV disease is imperative. In recent years, the number of patients undergoing organ and tissue transplantations has increased markedly. HCMV is the most frequent cause of death in immunocompromised transplant recipients, thereby confirming the need for rapid and reliable laboratory diagnosis. Lymphocytes, monocytes, and possibly arterial endothelial or smooth muscle cells, are sites of HCMV latency. Therefore, prevention of HCMV infections in immunocompromised individuals (*e.g.*, transplant recipients) includes use of HCMV-negative blood products and organs. Additionally, HCMV can be spread transplacentally, and to newborns by contact with infected cervical secretions during birth. Thus, a rapid, sensitive, and specific assay for detecting HCMV in body fluids or secretions may be desirable as a means to monitor infection, and consequently, determine the necessity of cesarean section.

Diagnosis of HCMV infection may be performed by conventional cell culture using human fibroblasts; shell vial centrifugation culture utilizing monoclonal antibodies and immunofluorescent staining techniques; serological methods; the HCMV antigenemia assay which employs a monoclonal antibody to detect HCMV antigen in peripheral blood leukocytes (PBLs); or by nucleic acid hybridization assays. These various methods have their advantages and limitations. Conventional cell culture is sensitive but slow, as cytopathic effect (CPE) may take 30 or more days to develop. Shell vial centrifugation is more rapid but still requires 24-48 hours for initial results. Both culture methods are affected by antiviral therapy. In immunocompromised patients, the ability to mount IgG and/or IgM antibody responses to HCMV infection are impaired, and serological methods are thus not reliable in this setting. Alternatively, IgM antibodies may be persistent for months after infection is resolved, and thus their presence may not be indicative of active infection. The HCMV antigenemia assay is labor intensive and is not applicable to specimens other than PBLs.

Recent advances in molecular biology have spurred the use of DNA probes in attempts to provide a more rapid, sensitive and specific assay for detecting HCMV in clinical specimens. For example, radiolabeled DNA probes have been used to hybridize to tissue cultures infected with or by HCMV, or in clinical samples suspected of containing HCMV ("hybridization assays"). However, probing of tissue cultures requires at least 18-24 hours for growth to amplify the antigen (HCMV) to be detected, if present, and additional time for development of autoradiographic detection systems. Using hybridization assays for assaying clinical specimens for HCMV may lack sensitivity, depending upon the titer of virus and the clinical sample assayed. Detection of HCMV in clinical samples has been reported using the

polymerase chain reaction (PCR) to enzymatically amplify HCMV DNA. Methods using PCR compare favorably with virus isolation, *in situ* hybridization assays, and Southern blotting; *See, e.g.*, Bamborschke *et al*,. J. Neurol., 239:205 [1992]; Drouet *et al.*, J. Virol. Meth., 45:259 [1993]; Einsele *et al.*, Blood 77:1104-1110 [1991]; Einsele *et al.*, Lancet 338:1170 [1991]; Lee *et al.*, Aust. NZ J. Med., 22:249 [1992]; Miller *et al.*, J. Clin. Microbiol., 32:5 [1994]; Rowley *et al.*, Transplant. 51:1028 [1991]; Spector *et al.* J. Clin. Microbiol., 30:2359 [1992]; and Stanier *et al.*, Mol. Cell. Probes 8:51 [1992]). Others, comparing the HCMV antigenemia assay with PCR methods, have found PCR methods as efficient or slightly more efficient in the detection of HCMV (van Dorp *et al.* (1992) Transplant. 54:661; Gerna *et al.* (1991) J. Infect. Dis. 164:488; Vleiger *et al.* (1992) Bone Marrow Transplant. 9:247; Zipeto *et al.* (1992) J. Clin. Microbiol. 30:527]. In addition, PCR methods have exhibited great sensitivity when specimens other than PBLs are assayed (Natori *et al.*, Kansenshogaku Zasshi 67:1011 [1993]; Peterson *et al.*, Medicine 59:283 [1980]; Prosch *et al.*, J. Med. Virol., 38:246 [1992]; Ratnamohan *et al.*, J. Med. Virol. 38:252 [1992]). However, because of the dangers of false positive reactions, these PCR-based procedures require rigid controls to prevent contamination and carry over (Ehrlich *et al.*, in *PCR-Based Diagnostics in Infectious Diseases*, Ehrlich and Greenberg (eds), Blackwell Scientific Publications, [1994], pp.3-18). Therefore, there exists a need for a rapid, sensitive, and specific assay for HCMV that has a reduced risk of false positive result due to contamination by reaction product carried over from other samples.

As shown herein, the Invader™-directed cleavage assay is rapid, sensitive and specific. Because the accumulated products do not contribute to the further accumulation of signal, reaction products carried over from one standard (*i.e.*, non-sequential) Invader™-directed cleavage assay to another cannot promote false positive results. The use of multiple sequential Invader™-directed cleavage assays will further boost the sensitivity of HCMV detection without sacrifice of these advantages.

XI. Effect of Arrestor™ Oligonucleotides on Signal and Background in Sequential Invasive Cleavage Reactions

As described above, and demonstrated in Example 36, the concentration of the probe that is cleaved can be used to increase the rate of signal accumulation, with higher concentrations of probe yielding higher final signal. However, the presence of large amounts of residual uncleaved probe can present problems for subsequent use of the cleaved products

for detection or for further amplification. If the subsequent step is a simple detection (*e.g.*, by gel resolution), the excess uncut material may cause background by streaking or scattering of signal, or by overwhelming a detector (*e.g.*, over-exposing a film in the case of radioactivity, or exceeding the quantitative detection limits of a fluorescence imager). This can be overcome by partitioning the product from the uncut probe (*e.g.*, by using the charge reversal method described in Example 22).

In more complex detection methods, the cleaved product may be intended to interact with another entity to indicate cleavage. As noted above, the cleaved product can be used in any reaction that makes use of oligonucleotides, such as hybridization, primer extension, ligation, or the direction of invasive cleavage. In each of these cases, the fate of the residual uncut probe must be considered in the design of the reaction. In a primer extension reaction, the uncut probe can hybridize to a template for extension. If cleavage is required to reveal the correct 3' end for extension, the hybridized uncut probe will not be extended. It will, however, compete with the cleaved product for the template. If the template is in excess of the combination of cleaved and uncleaved probe, then both of the latter should be able to find a copy of template for binding. If, however, the template is limiting, the competition reduces the portion of the cleaved probe that can find successfully bind to the available template. If a vast excess of probe was used to drive the initial reaction, the remainder may also be in vast excess over the cleavage product, and thus may provide a very effective competitor, thereby reducing the amount of the final reaction (*e.g.*, extension) product for ultimate detection.

The participation of the uncut probe material in a secondary reaction can also contribute to background in these reactions. While the presentation of a cleaved probe for a subsequent reaction may represent an ideal substrate for the enzyme to be used in the next step, some enzymes may also be able to act, albeit inefficiently, on the uncut probe as well. It was shown in Example 43 that transcription can be promoted from a nicked promoter even when one side of the nick has additional unpaired nucleotides (termed a "branched promoter" in this Example). Similarly, when the subsequent reaction is to be an invasive cleavage structure, the uncleaved probe may bind to the elements intended to form the second cleavage structure with the cleaved probe. Two of the possible configurations are shown schematically in Figs. 105 and 106. The right hand structure in the second step in each Figure shows a possible configuration formed by the secondary reaction elements (*e.g.*, secondary targets and/or probes) and the uncleaved primary probe. In each of these cases, it was found that the 5' nucleases described herein can direct some measure of cleavage of these defective structures.

Even at a low level, this aberrant cleavage can be misinterpreted as positive target-specific cleavage signal.

With these negative effects of the surfeit of uncut probe considered, there is clearly a need for some method of preventing these interactions. As noted above, it is possible to partition the cleaved product from the uncut probe after the primary reaction by traditional methods. However, these methods are often time consuming, may be expensive (*e.g.*, disposable columns, gels, etc.), and may increase the risk for sample mishandling or contamination. It is far preferable to configure the sequential reactions such that the original sample need not be removed to a new vessel for subsequent reaction.

The present invention provides a method for reducing the interactions between the primary probe and any subsequent reactants. This method provides a means of specifically diverting the uncleaved probes from participation in the subsequent reactions. The diversion is accomplished by the inclusion in the next reaction step an agent designed to specifically interact with the uncleaved primary probe. While the primary probe in an invasive cleavage reaction is discussed for reasons of convenience, it is contemplated that the Arrestors™ may be used at any reaction step within a chain of invasive cleavage steps, as needed or desired for the design of an assay. It is not intended that the Arrestors™ of the present invention be limited to any particular step.

The method of diverting the residual uncut probes from a primary reaction makes use of agents that can be specifically designed or selected to bind to the uncleaved probe molecules with greater affinity than to the cleaved probes, thereby allowing the cleaved probe species to effectively compete for the elements of the subsequent reaction, even when the uncut probe is present in vast excess. These agents have been termed "Arrestors™," due to their function of stopping or arresting the primary probe from participation in the later reaction. In various Examples below, an oligonucleotide is provided as an Arrestor™ in an invasive cleavage assay. It can be appreciated that any molecule or chemical that can discriminate between the full-length uncut probe and the cleaved probe, and that can bind or otherwise disable the uncleaved probe preferentially may be configured to act as an Arrestor™ within the meaning of the present invention. For example, antibodies can be derived with such specificity, as can the "aptamers" that can be selected through multiple steps of *in vitro* amplification (*e.g.*, "SELEX," U.S. Patent Nos. 5,270,163 and 5,567,588; herein incorporated by reference) and specific rounds of capture or other selection means.

In one embodiment, the Arrestor™ is an oligonucleotide. In another embodiment the oligonucletide Arrestor™ is a composite oligonucleotide, comprising two or more short olignucleotides that are not covalently linked, but that bind cooperatively and are stabilized by co-axial stacking. In a preferred embodiment, the oligonucleotide is modified to reduce interactions with the cleavage agents of the present invention. When an oligonucleotide is used as an Arrestor™, it is intended that it not participate in the subsequent reactive step. Consideration of the schematic diagrams in Figs. 105 and 106, particularly the right-most Figure in step 2b of each Figure, will show that the binding of the Arrestor™ to the primary probe may, either with the participation of the secondary target, or without such participation, create a bifurcated structure that is a substrate for cleavage by the 5' nucleases used in some embodiments of the methods of the present invention. Formation of such structures would lead to some level of unintended cleavage that could contribute to background, reduce specific signal or compete for the enzyme. It is preferable to provide Arrestors™ that will not create such cleavage structures. One method of doing this is to add to the Arrestor™ such modifications as have been found to reduce the activity of Invader™ oligonucleotides, as the Invader™ oligonucleotides occupy a similar position within a cleavage structure (*i.e.*, the 3' end of the Invader™ oligonucleotide positions the site of cleavage of an unpaired 5' arm). Modification of the 3' end of the Invader™ oligonucleotides was examined for the effects on cleavage in Example 35; a number of the modifications tested were found to be significantly debilitating to the function of the Invader™ oligonucleotide. Other modifications not described herein may be easily characterized by performing such a test using the cleavage enzyme to be used in the reaction for which the Arrestor™ is intended.

In a preferred embodiment, the backbone of an oligonucleotide Arrestor™ is modified. This may be done to increase the resistance to degradation by nucleases or temperature, or to provide duplex structure that is a less favorable substrate for the enzyme to be used (*e.g.*, A-form duplex vs. B-form duplex). In particularly preferred embodiment, the backbone modified oligonucleotide further comprises a 3' terminal modification. In a preferred embodiment, the modifications comprise 2' O-methyl substitution of the nucleic acid backbone, while in a particularly preferred embodiment, the 2' O-methyl modified oligonucleotide further comprises a 3' terminal amine group.

The purpose of the Arrestor™ is to allow the minority population of cleaved probe to effectively compete with the uncleaved probe for binding whatever elements are to be used in the next step. While an Arrestor™ that can discriminate between the two probe species

absolutely (*i.e.*, binding only to uncut and never to cut) may be of the greatest benefit in some embodiments, it is envisioned that in many applications, including the sequential Invader™ assays described herein, the Arrestors™ of the present invention may perform the intended function with only partial discrimination. When the Arrestor™ has some interaction with the cleaved probe, it may prevent detection of some portion of these cleavage products, thereby reducing the absolute level of signal generated from a given amount of target material. If this same Arrestor™ has the simultaneous effect of reducing the background of the reaction (*i.e.*, from non-target specific cleavage) by a factor that is greater than the factor of reduction in the specific signal, then the significance of the signal (*i.e.*, the ratio of signal to background), is increased, even with the lower amount of absolute signal. Any potential Arrestor™ design may be tested in a simple fashion by comparing the levels of background and specific signals from reactions that lack Arrestors™ to the levels of background and specific signal from similar reactions that include Arrestors™. Each of the reactions described in Examples 49-53 demonstrate the use of such comparisons, and these can easily be adapted by those skilled in the art to other Arrestor™ and target embodiments. What constitutes an acceptable level of tradeoff of absolute signal for specificity will vary for different applications (*e.g.*, target levels, read-out sensitivity, etc.), and can be determined by any individual user using the methods of the present invention.

EXPERIMENTAL

The following examples serve to illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

In the disclosure which follows, the following abbreviations apply: Afu (*Archaeoglobus fulgidus*); Mth (*Methanobacterium thermoautotrophicum*); Mja (*Methanococcus jannaschii*); Pfu (*Pyrococcus furiosus*); Pwo (*Pyrococcus woesei*); Taq (*Thermus aquaticus*); *Taq* DNAP, DNAP*Taq*, and *Taq* Pol I (*T. aquaticus* DNA polymerase I); DNAPStf (the Stoffel fragment of DNAP*Taq*); DNAPEc1 (*E. coli* DNA polymerase I); Tth (*Thermus thermophilus*); Ex. (Example); Fig. (Figure);°C (degrees Centigrade); *g* (gravitational field); hr (hour); min (minute); olio (oligonucleotide); rxn (reaction); vol (volume); w/v (weight to volume); v/v (volume to volume); BSA (bovine serum albumin); CTAB (cetyltrimethylammonium bromide); HPLC (high pressure liquid chromatography); DNA (deoxyribonucleic acid); p (plasmid); μl (microliters); ml (milliliters); μg (micrograms); mg (milligrams); M (molar); mM (milliMolar); μM (microMolar); pmoles (picomoles);

amoles (attomoles); zmoles (zeptomoles); nm (nanometers); kdal (kilodaltons); OD (optical density); EDTA (ethylene diamine tetra-acetic acid); FITC (fluorescein isothiocyanate); SDS (sodium dodecyl sulfate); $NaPO_4$ (sodium phosphate); NP-40 (Nonidet P-40); Tris (tris(hydroxymethyl)-aminomethane); PMSF (phenylmethylsulfonylfluoride); TBE (Tris-Borate-EDTA, *i.e.*, Tris buffer titrated with boric acid rather than HCl and containing EDTA); PBS (phosphate buffered saline); PPBS (phosphate buffered saline containing 1 mM PMSF); PAGE (polyacrylamide gel electrophoresis); Tween (polyoxyethylene-sorbitan); ATCC (American Type Culture Collection, Rockville, MD); DSMZ (Deutsche Sammlung von Mikroorganismen und Zellculturen, Braunschweig, Germany); Ambion (Ambion, Inc., Austin, TX); Boehringer (Boehringer Mannheim Biochemical, Indianapolis, IN); MJ Research (MJ Research, Watertown, MA; Sigma (Sigma Chemical Company, St. Louis, MO); Dynal (Dynal A.S., Oslo, Norway); Gull (Gull Laboratories, Salt Lake City, UT); Epicentre (Epicentre Technologies, Madison, WI); MJ Research (MJ Research, Watertown,MA); National Biosciences (National Biosciences, Plymouth, MN); NEB (New England Biolabs, Beverly, MA); Novagen (Novagen, Inc., Madison, WI); Perkin Elmer (Perkin-Elmer/ABI, Norwalk, CT); Promega (Promega, Corp., Madison, WI); Stratagene (Stratagene Cloning Systems, La Jolla, CA); Clonetech (Clonetech, Palo Alto, CA) Pharmacia (Pharmacia, Piscataway, NJ); Milton Roy (Milton Roy, Rochester, NY); Amersham (Amersham International, Chicago, IL); and USB (U.S. Biochemical, Cleveland, OH).

EXAMPLE 1

Characteristics Of Native Thermostable DNA Polymerases

A. 5' Nuclease Activity Of DNAP*Taq*

During the polymerase chain reaction (PCR) (Saiki et al., Science 239:487 [1988]; Mullis and Faloona, Meth. Enzymol., 155:335 [1987]), DNAP*Taq* is able to amplify many, but not all, DNA sequences. One sequence that cannot be amplified using DNAP*Taq* is shown in Fig. 5 (Hairpin structure is SEQ ID NO:15, Fig. 5 also shows a primer: SEQ ID NO:17.) This DNA sequence has the distinguishing characteristic of being able to fold on itself to form a hairpin with two single-stranded arms, which correspond to the primers used in PCR.

To test whether this failure to amplify is due to the 5' nuclease activity of the enzyme, the abilities of DNAP*Taq* and DNAPStf to amplify this DNA sequence during 30 cycles of PCR were compared. Synthetic oligonucleotides were obtained from The Biotechnology

Center at the University of Wisconsin-Madison. The DNAP*Taq* and DNAPStf were from Perkin Elmer (*i.e.*, Amplitaq™ DNA polymerase and the Stoffel fragment of Amplitaq™ DNA polymerase). The substrate DNA comprised the hairpin structure shown in Fig. 6 cloned in a double-stranded form into pUC19. The primers used in the amplification are listed as SEQ ID NOS:16-17. Primer SEQ ID NO:17 is shown annealed to the 3' arm of the hairpin structure in Fig. 5. Primer SEQ ID NO:16 is shown as the first 20 nucleotides in bold on the 5' arm of the hairpin in Fig. 5.

Polymerase chain reactions comprised 1 ng of supercoiled plasmid target DNA, 5 pmoles of each primer, 40 μM each dNTP, and 2.5 units of DNAP*Taq* or DNAPStf, in a 50 μl solution of 10 mM Tris•Cl pH 8.3. The DNAP*Taq* reactions included 50 mM KCl and 1.5 mM $MgCl_2$. The temperature profile was 95°C for 30 sec., 55°C for 1 min. and 72°C for 1 min., through 30 cycles. Ten percent of each reaction was analyzed by gel electrophoresis through 6% polyacrylamide (cross-linked 29:1) in a buffer of 45 mM Tris•Borate, pH 8.3, 1.4 mM EDTA.

The results are shown in Fig. 6. The expected product was made by DNAPStf (indicated simply as "S") but not by DNAP*Taq* (indicated as "T"). It was concluded that the 5' nuclease activity of DNAP*Taq* is responsible for the lack of amplification of this DNA sequence.

To test whether the 5' unpaired nucleotides in the substrate region of this structured DNA are removed by DNAP*Taq*, the fate of the end-labeled 5' arm during four cycles of PCR was compared using the same two polymerases (Fig. 7). The hairpin templates, such as the one described in Fig. 5, were made using DNAPStf and a ^{32}P-5'-end-labeled primer. The 5'-end of the DNA was released as a few large fragments by DNAP*Taq* but not by DNAPStf. The sizes of these fragments (based on their mobilities) show that they contain most or all of the unpaired 5' arm of the DNA. Thus, cleavage occurs at or near the base of the bifurcated duplex. These released fragments terminate with 3' OH groups, as evidenced by direct sequence analysis, and the abilities of the fragments to be extended by terminal deoxynucleotidyl transferase.

Figs. 8-10 show the results of experiments designed to characterize the cleavage reaction catalyzed by DNAP*Taq*. Unless otherwise specified, the cleavage reactions comprised 0.01 pmoles of heat-denatured, end-labeled hairpin DNA (with the unlabeled complementary strand also present), 1 pmole primer (complementary to the 3' arm) and 0.5 units of DNAP*Taq* (estimated to be 0.026 pmoles) in a total volume of 10μl of 10 mM Tris-

Cl, ph 8.5, 50 mM KCl and 1.5 mM $MgCl_2$. As indicated, some reactions had different concentrations of KCl, and the precise times and temperatures used in each experiment are indicated in the individual Figures. The reactions that included a primer used the one shown in Fig. 5 (SEQ ID NO:17). In some instances, the primer was extended to the junction site by providing polymerase and selected nucleotides.

Reactions were initiated at the final reaction temperature by the addition of either the $MgCl_2$ or enzyme. Reactions were stopped at their incubation temperatures by the addition of 8 μl of 95% formamide with 20 mM EDTA and 0.05% marker dyes. The T_m calculations listed were made using the Oligo™ primer analysis software from National Biosciences, Inc. These were determined using 0.25 μM as the DNA concentration, at either 15 or 65 mM total salt (the 1.5 mM $MgCl_2$ in all reactions was given the value of 15 mM salt for these calculations).

Fig. 8 is an autoradiogram containing the results of a set of experiments and conditions on the cleavage site. Fig. 8A is a determination of reaction components that enable cleavage. Incubation of 5'-end-labeled hairpin DNA was for 30 minutes at 55°C, with the indicated components. The products were resolved by denaturing polyacrylamide gel electrophoresis and the lengths of the products, in nucleotides, are indicated. Fig. 8B describes the effect of temperature on the site of cleavage in the absence of added primer. Reactions were incubated in the absence of KCl for 10 minutes at the indicated temperatures. The lengths of the products, in nucleotides, are indicated.

Surprisingly, cleavage by DNAP*Taq* requires neither a primer nor dNTPs (*See* Fig. 8A). Thus, the 5' nuclease activity can be uncoupled from polymerization. Nuclease activity requires magnesium ions, though manganese ions can be substituted, albeit with potential changes in specificity and activity. Neither zinc nor calcium ions support the cleavage reaction. The reaction occurs over a broad temperature range, from 25°C to 85°C, with the rate of cleavage increasing at higher temperatures.

Still referring to Fig. 8, the primer is not elongated in the absence of added dNTPs. However, the primer influences both the site and the rate of cleavage of the hairpin. The change in the site of cleavage (Fig. 8A) apparently results from disruption of a short duplex formed between the arms of the DNA substrate. In the absence of primer, the sequences indicated by underlining in Fig. 5 could pair, forming an extended duplex. Cleavage at the end of the extended duplex would release the 11 nucleotide fragment seen on the Fig. 8A lanes with no added primer. Addition of excess primer (Fig. 8A, lanes 3 and 4) or incubation

at an elevated temperature (Fig. 8B) disrupts the short extension of the duplex and results in a longer 5' arm and, hence, longer cleavage products.

The location of the 3' end of the primer can influence the precise site of cleavage. Electrophoretic analysis revealed that in the absence of primer (Fig. 8B), cleavage occurs at the end of the substrate duplex (either the extended or shortened form, depending on the temperature) between the first and second base pairs. When the primer extends up to the base of the duplex, cleavage also occurs one nucleotide into the duplex. However, when a gap of four or six nucleotides exists between the 3' end of the primer and the substrate duplex, the cleavage site is shifted four to six nucleotides in the 5' direction.

Fig. 9 describes the kinetics of cleavage in the presence (Fig. 9A) or absence (Fig. 9B) of a primer oligonucleotide. The reactions were run at 55°C with either 50 mM KCl (Fig. 9A) or 20 mM KCl (Fig. 9B). The reaction products were resolved by denaturing polyacrylamide gel electrophoresis and the lengths of the products, in nucleotides, are indicated. "M", indicating a marker, is a 5' end-labeled 19-nt oligonucleotide. Under these salt conditions, Figs. 9A and 9B indicate that the reaction appears to be about twenty times faster in the presence of primer than in the absence of primer. This effect on the efficiency may be attributable to proper alignment and stabilization of the enzyme on the substrate.

The relative influence of primer on cleavage rates becomes much greater when both reactions are run in 50 mM KCl. In the presence of primer, the rate of cleavage increases with KCl concentration, up to about 50 mM. However, inhibition of this reaction in the presence of primer is apparent at 100 mM and is complete at 150 mM KCl. In contrast, in the absence of primer the rate is enhanced by concentration of KCl up to 20 mM, but it is reduced at concentrations above 30 mM. At 50 mM KCl, the reaction is almost completely inhibited. The inhibition of cleavage by KCl in the absence of primer is affected by temperature, being more pronounced at lower temperatures.

Recognition of the 5' end of the arm to be cut appears to be an important feature of substrate recognition. Substrates that lack a free 5' end, such as circular M13 DNA, cannot be cleaved under any conditions tested. Even with substrates having defined 5' arms, the rate of cleavage by DNAP*Taq* is influenced by the length of the arm. In the presence of primer and 50 mM KCl, cleavage of a 5' extension that is 27 nucleotides long is essentially complete within 2 minutes at 55°C. In contrast, cleavages of molecules with 5' arms of 84 and 188 nucleotides are only about 90% and 40% complete after 20 minutes. Incubation at higher temperatures reduces the inhibitory effects of long extensions indicating that secondary

structure in the 5' arm or a heat-labile structure in the enzyme may inhibit the reaction. A mixing experiment, run under conditions of substrate excess, shows that the molecules with long arms do not preferentially tie up the available enzyme in non-productive complexes. These results may indicate that the 5' nuclease domain gains access to the cleavage site at the end of the bifurcated duplex by moving down the 5' arm from one end to the other. Longer 5' arms would be expected to have more adventitious secondary structures (particularly when KCl concentrations are high), which would be likely to impede this movement.

Cleavage does not appear to be inhibited by long 3' arms of either the substrate strand target molecule or pilot nucleic acid, at least up to 2 kilobases. At the other extreme, 3' arms of the pilot nucleic acid as short as one nucleotide can support cleavage in a primer-independent reaction, albeit inefficiently. Fully paired oligonucleotides do not elicit cleavage of DNA templates during primer extension.

The ability of DNAP*Taq* to cleave molecules even when the complementary strand contains only one unpaired 3' nucleotide may be useful in optimizing allele-specific PCR. PCR primers that have unpaired 3' ends could act as pilot oligonucleotides to direct selective cleavage of unwanted templates during preincubation of potential template-primer complexes with DNAP*Taq* in the absence of nucleoside triphosphates.

B. 5' Nuclease Activities Of Other DNAPs

To determine whether other 5' nucleases in other DNAPs would be suitable for the present invention, an array of enzymes, several of which were reported in the literature to be free of apparent 5' nuclease activity, were examined. The ability of these other enzymes to cleave nucleic acids in a structure-specific manner was tested using the hairpin substrate shown in Fig. 5 under conditions reported to be optimal for synthesis by each enzyme.

DNAPEcl and DNAP Klenow were obtained from Promega; the DNAP of *Pyrococcus furious* ("Pfu", Bargseid *et al.*, Strategies 4:34 [1991]) was from Stratagene; the DNAP of *Thermococcus litoralis* ("Tli", Vent™(exo-), Perler *et al.*, Proc. Natl. Acad. Sci. USA 89:5577 [1992] was from New England Biolabs; the DNAP of *Thermus flavus* ("Tfl", Kaledin *et al.*, Biokhimiya 46:1576 [1981] was from Epicentre Technologies; and the DNAP of *Thermus thermophilus* ("Tth", Carballeira *et al.*, Biotechn., 9:276 [1990]; Myers *et al.*, Biochem., 30:7661 (1991)] was from U.S. Biochemicals.

0.5 units of each DNA polymerase was assayed in a 20 μl reaction, using either the buffers supplied by the manufacturers for the primer-dependent reactions, or 10 mM Tris•Cl,

pH 8.5, 1.5 mM $MgCl_2$, and 20mM KCl. Reaction mixtures were at held 72°C before the addition of enzyme.

Fig. 10 is an autoradiogram recording the results of these tests. Fig. 10A demonstrates reactions of endonucleases of DNAPs of several thermophilic bacteria. The reactions were incubated at 55°C for 10 minutes in the presence of primer or at 72°C for 30 minutes in the absence of primer, and the products were resolved by denaturing polyacrylamide gel electrophoresis. The lengths of the products, in nucleotides, are indicated. Fig. 10B demonstrates endonucleolytic cleavage by the 5' nuclease of DNAPEcl. The DNAPEcl and DNAP Klenow reactions were incubated for 5 minutes at 37°C. Note the light band of cleavage products of 25 and 11 nucleotides in the DNAPEcl lanes (made in the presence and absence of primer, respectively). Fig. 8A also demonstrates DNAP*Taq* reactions in the presence (+) or absence (-) of primer. These reactions were run in 50 mM and 20 mM KCl, respectively, and were incubated at 55°C for 10 minutes.

Referring to Fig. 10A, DNAPs from the eubacteria *Thermus thermophilus* and *Thermus flavus* cleave the substrate at the same place as DNAP*Taq*, both in the presence and absence of primer. In contrast, DNAPs from the archaebacteria *Pyrococcus furiosus* and *Thermococcus litoralis* are unable to cleave the substrates endonucleolytically. The DNAPs from *Pyrococcus furious* and *Thermococcus litoralis* share little sequence homology with eubacterial enzymes (Ito *et al.*, *Nucl. Acids Res.* 19:4045 (1991); Mathur *et al.*, *Nucl. Acids. Res.* 19:6952 (1991); *see also* Perler *et al.*). Referring to Fig. 10B, DNAPEcl also cleaves the substrate, but the resulting cleavage products are difficult to detect unless the 3' exonuclease is inhibited. The amino acid sequences of the 5' nuclease domains of DNAPEcl and DNAP*Taq* are about 38% homologous (Gelfand, *supra*).

The 5' nuclease domain of DNAP*Taq* also shares about 19% homology with the 5' exonuclease encoded by gene 6 of bacteriophage T7 (Dunn *et al.*, J. Mol. Biol.,166:477 [1983]). This nuclease, which is not covalently attached to a DNAP polymerization domain, is also able to cleave DNA endonucleolytically, at a site similar or identical to the site that is cut by the 5' nucleases described above, in the absence of added primers.

C. Transcleavage

The ability of a 5' nuclease to be directed to cleave efficiently at any specific sequence was demonstrated in the following experiment. A partially complementary oligonucleotide termed a "pilot oligonucleotide" was hybridized to sequences at the desired

point of cleavage. The non-complementary part of the pilot oligonucleotide provided a structure analogous to the 3' arm of the template (*see* Fig. 5), whereas the 5' region of the substrate strand became the 5' arm. A primer was provided by designing the 3' region of the pilot so that it would fold on itself creating a short hairpin with a stabilizing tetra-loop (Antao *et al.*, *Nucl. Acids Res.* 19:5901 [1991). Two pilot oligonucleotides are shown in Fig. 11A. Oligonucleotides 19-12 (SEQ ID NO:18), 30-12 (SEQ ID NO:19) and 30-0 (SEQ ID NO:20) are 31, 42 or 30 nucleotides long, respectively. However, oligonucleotides 19-12 (SEQ ID NO:18) and 34-19 (SEQ ID NO:19) have only 19 and 30 nucleotides, respectively, that are complementary to different sequences in the substrate strand. The pilot oligonucleotides are calculated to melt off their complements at about 50°C (19-12) and about 75°C (30-12). Both pilots have 12 nucleotides at their 3' ends, which act as 3' arms with base-paired primers attached.

To demonstrate that cleavage could be directed by a pilot oligonucleotide, a single-stranded target DNA with DNAP*Taq* was incubated in the presence of two potential pilot oligonucleotides. The transcleavage reactions, where the target and pilot nucleic acids are not covalently linked, includes 0.01 pmoles of single end-labeled substrate DNA, 1 unit of DNAP*Taq* and 5 pmoles of pilot oligonucleotide in a volume of 20 μl of the same buffers. These components were combined during a one minute incubation at 95°C, to denature the PCR-generated double-stranded substrate DNA, and the temperatures of the reactions were then reduced to their final incubation temperatures. Oligonucleotides 30-12 and 19-12 can hybridize to regions of the substrate DNAs that are 85 and 27 nucleotides from the 5' end of the targeted strand.

Fig. 19 shows the complete 206-mer sequence (SEQ ID NO:27). The 206-mer was generated by PCR . The M13/pUC 24-mer reverse sequencing (-48) primer and the M13/pUC sequencing (-47) primer from NEB (catalogue nos. 1233 and 1224 respectively) were used (50 pmoles each) with the pGEM3z(f+) plasmid vector (Promega) as template (10 ng) containing the target sequences. The conditions for PCR were as follows: 50 μM of each dNTP and 2.5 units of Taq DNA polymerase in 100 μl of 20 mM Tris-Cl, pH 8.3, 1.5 mM $MgCl_2$, 50 mM KCl with 0.05% Tween-20 and 0.05% NP-40. Reactions were cycled 35 times through 95°C for 45 seconds, 63°C for 45 seconds, then 72°C for 75 seconds. After cycling, reactions were finished off with an incubation at 72°C for 5 minutes. The resulting fragment was purified by electrophoresis through a 6% polyacrylamide gel (29:1 cross link) in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA, visualized by ethidium bromide

staining or autoradiography, excised from the gel, eluted by passive diffusion, and concentrated by ethanol precipitation.

Cleavage of the substrate DNA occurred in the presence of the pilot oligonucleotide 19-12 at 50°C (Fig. 11B, lanes 1 and 7) but not at 75°C (lanes 4 and 10). In the presence of oligonucleotide 30-12 cleavage was observed at both temperatures. Cleavage did not occur in the absence of added oligonucleotides (lanes 3, 6 and 12) or at about 80°C even though at 50°C adventitious structures in the substrate allowed primer-independent cleavage in the absence of KCl (Fig. 11B, lane 9). A non-specific oligonucleotide with no complementarity to the substrate DNA did not direct cleavage at 50°C, either in the absence or presence of 50 mM KCl (lanes 13 and 14). Thus, the specificity of the cleavage reactions can be controlled by the extent of complementarity to the substrate and by the conditions of incubation.

D. Cleavage Of RNA

A shortened RNA version of the sequence used in the transcleavage experiments discussed above was tested for its ability to serve as a substrate in the reaction. The RNA is cleaved at the expected place, in a reaction that is dependent upon the presence of the pilot oligonucleotide. The RNA substrate, made by T7 RNA polymerase in the presence of (α-^{32}P)UTP, corresponds to a truncated version of the DNA substrate used in Fig. 11B. Reaction conditions were similar to those in used for the DNA substrates described above, with 50 mM KCl; incubation was for 40 minutes at 55°C. The pilot oligonucleotide used is termed 30-0 (SEQ ID NO:20) and is shown in Fig. 12A.

The results of the cleavage reaction is shown in Fig. 13B. The reaction was run either in the presence or absence of DNAP*Taq* or pilot oligonucleotide as indicated in Fig. 12B.

Strikingly, in the case of RNA cleavage, a 3' arm is not required for the pilot oligonucleotide. It is very unlikely that this cleavage is due to previously described RNaseH, which would be expected to cut the RNA in several places along the 30 base-pair long RNA-DNA duplex. The 5' nuclease of DNAP*Taq* is a structure-specific RNaseH that cleaves the RNA at a single site near the 5' end of the heteroduplexed region.

It is surprising that an oligonucleotide lacking a 3' arm is able to act as a pilot in directing efficient cleavage of an RNA target because such oligonucleotides are unable to direct efficient cleavage of DNA targets using native DNAPs. However, some 5' nucleases of the present invention (for example, clones E, F and G of Fig. 4) can cleave DNA in the absence of a 3' arm. In other words, a non-extendable cleavage structure is not required for

specific cleavage with some 5' nucleases of the present invention derived from thermostable DNA polymerases.

Tests were then conducted to determine whether cleavage of an RNA template by DNAP*Taq* in the presence of a fully complementary primer could help explain why DNAP*Taq* is unable to extend a DNA oligonucleotide on an RNA template, in a reaction resembling that of reverse transcriptase. Another thermophilic DNAP, DNAPTth, is able to use RNA as a template, but only in the presence of Mn++, so it was predicted that this enzyme would not cleave RNA in the presence of this cation. Accordingly, an RNA molecule was incubated with an appropriate pilot oligonucleotide in the presence of DNAP*Taq* or DNAPTth, in buffer containing either Mg++ or Mn++. As expected, both enzymes cleaved the RNA in the presence of Mg++. However, DNAP*Taq*, but not DNAPTth, degraded the RNA in the presence of Mn++. It was concluded that the 5' nuclease activities of many DNAPs may contribute to their inability to use RNA as templates.

EXAMPLE 2

Generation Of 5' Nucleases From Thermostable DNA Polymerases

Thermostable DNA polymerases were generated which have reduced synthetic activity, an activity that is an undesirable side-reaction during DNA cleavage in the detection assay of the invention, yet have maintained thermostable nuclease activity. The result is a thermostable polymerase which cleaves nucleic acids DNA with extreme specificity.

Type A DNA polymerases from eubacteria of the genus *Thermus* share extensive protein sequence identity (90% in the polymerization domain, using the Lipman-Pearson method in the DNA analysis software from DNAStar, WI) and behave similarly in both polymerization and nuclease assays. Therefore, the genes for the DNA polymerase of *Thermus aquaticus* (DNAP*Taq*) and *Thermus flavus* (DNAPTfl) are used as representatives of this class. Polymerase genes from other eubacterial organisms, such as *Thermus thermophilus*, *Thermus sp.*, *Thermotoga maritima*, *Thermosipho africanus* and *Bacillus stearothermophilus* are equally suitable. The DNA polymerases from these thermophilic organisms are capable of surviving and performing at elevated temperatures, and can thus be used in reactions in which temperature is used as a selection against non-specific hybridization of nucleic acid strands.

The restriction sites used for deletion mutagenesis, described below, were chosen for convenience. Different sites situated with similar convenience are available in the *Thermus*

thermophilus gene and can be used to make similar constructs with other Type A polymerase genes from related organisms.

A. Creation Of 5' Nuclease Constructs

1. Modified DNAP*Taq* Genes

The first step was to place a modified gene for the *Taq* DNA polymerase on a plasmid under control of an inducible promoter. The modified *Taq* polymerase gene was isolated as follows: The *Taq* DNA polymerase gene was amplified by polymerase chain reaction from genomic DNA from *Thermus aquaticus*, strain YT-1 (Lawyer *et al.*, *supra*), using as primers the oligonucleotides described in SEQ ID NOS:13-14. The resulting fragment of DNA has a recognition sequence for the restriction endonuclease EcoRI at the 5' end of the coding sequence and a *Bgl*II sequence at the 3' end. Cleavage with *Bgl*II leaves a 5' overhang or "sticky end" that is compatible with the end generated by *Bam*HI. The PCR-amplified DNA was digested with *Eco*RI and *Bam*HI. The 2512 bp fragment containing the coding region for the polymerase gene was gel purified and then ligated into a plasmid which contains an inducible promoter.

In one embodiment of the invention, the pTTQ18 vector, which contains the hybrid *trp-lac* (*tac*) promoter, was used (Stark, *Gene* 5:255 [1987]) and shown in Fig. 13. The *tac* promoter is under the control of the *E. coli lac* repressor. Repression allows the synthesis of the gene product to be suppressed until the desired level of bacterial growth has been achieved, at which point repression is removed by addition of a specific inducer, isopropyl-β-D-thiogalactopyranoside (IPTG). Such a system allows the expression of foreign proteins that may slow or prevent growth of transformants.

Bacterial promoters, such as *tac*, may not be adequately suppressed when they are present on a multiple copy plasmid. If a highly toxic protein is placed under control of such a promoter, the small amount of expression leaking through can be harmful to the bacteria. In another embodiment of the invention, another option for repressing synthesis of a cloned gene product was used. The non-bacterial promoter, from bacteriophage T7, found in the plasmid vector series pET-3 was used to express the cloned mutant *Taq* polymerase genes (Fig. 15; Studier and Moffatt, J. Mol. Biol., 189:113 [1986]). This promoter initiates transcription only by T7 RNA polymerase. In a suitable strain, such as BL21(DE3)pLYS, the gene for this RNA polymerase is carried on the bacterial genome under control of the *lac* operator. This arrangement has the advantage that expression of the multiple copy gene (on

the plasmid) is completely dependent on the expression of T7 RNA polymerase, which is easily suppressed because it is present in a single copy.

For ligation into the pTTQ18 vector (Fig. 13), the PCR product DNA containing the *Taq* polymerase coding region (mut*Taq*, clone 4B, SEQ ID NO:21) was digested with *Eco*RI and *Bgl*II and this fragment was ligated under standard "sticky end" conditions (Sambrook *et al. Molecular Cloning*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 1.63-1.69 [1989]) into the *Eco*RI and *Bam*HI sites of the plasmid vector pTTQ18. Expression of this construct yields a translational fusion product in which the first two residues of the native protein (Met-Arg) are replaced by three from the vector (Met-Asn-Ser), but the remainder of the natural protein would not change. The construct was transformed into the JM109 strain of *E. coli* and the transformants were plated under incompletely repressing conditions that do not permit growth of bacteria expressing the native protein. These plating conditions allow the isolation of genes containing pre-existing mutations, such as those that result from the infidelity of *Taq* polymerase during the amplification process.

Using this amplification/selection protocol, a clone (depicted in Fig. 3B) containing a mutated *Taq* polymerase gene (mut*Taq*, clone 3B) was isolated. The mutant was first detected by its phenotype, in which temperature-stable 5' nuclease activity in a crude cell extract was normal, but polymerization activity was almost absent (approximately less than 1% of wild type *Taq* polymerase activity).

DNA sequence analysis of the recombinant gene showed that it had changes in the polymerase domain resulting in two amino acid substitutions: an A to G change at nucleotide position 1394 causes a Glu to Gly change at amino acid position 465 (numbered according to the natural nucleic and amino acid sequences, SEQ ID NOS:1 and 4) and another A to G change at nucleotide position 2260 causes a Gln to Arg change at amino acid position 754. Because the Gln to Gly mutation is at a nonconserved position and because the Glu to Arg mutation alters an amino acid that is conserved in virtually all of the known Type A polymerases, this latter mutation is most likely the one responsible for curtailing the synthesis activity of this protein. The nucleotide sequence for the Fig. 3B construct is given in SEQ ID NO:21. The enzyme encoded by this sequence is referred to as Cleavase® A/G.

Subsequent derivatives of DNAP*Taq* constructs were made from the mut*Taq* gene, thus, they all bear these amino acid substitutions in addition to their other alterations, unless these particular regions were deleted. These mutated sites are indicated by black boxes at these locations in the diagrams in Fig. 3. In Fig. 3, the designation "3' Exo" is used to

indicate the location of the 3' exonuclease activity associated with Type A polymerases which is not present in DNAP*Taq*. All constructs except the genes shown in Figs. 3E, F and G were made in the pTTQ18 vector.

The cloning vector used for the genes in Figs. 3E and F was from the commercially available pET-3 series, described above. Though this vector series has only a BamHI site for cloning downstream of the T7 promoter, the series contains variants that allow cloning into any of the three reading frames. For cloning of the PCR product described above, the variant called pET-3c was used (Fig. 14). The vector was digested with *Bam*HI, dephosphorylated with calf intestinal phosphatase, and the sticky ends were filled in using the Klenow fragment of DNAPEc1 and dNTPs. The gene for the mutant *Taq* DNAP shown in Fig. 3B (mut*Taq*, clone 3B) was released from pTTQ18 by digestion with *Eco*RI and *Sal*I, and the "sticky ends" were filled in as was done with the vector. The fragment was ligated to the vector under standard blunt-end conditions (Sambrook *et al.*, *Molecular Cloning, supra*), the construct was transformed into the BL21(DE3)pLYS strain of *E. coli*, and isolates were screened to identify those that were ligated with the gene in the proper orientation relative to the promoter. This construction yields another translational fusion product, in which the first two amino acids of DNAP*Taq* (Met-Arg) are replaced by 13 from the vector plus two from the PCR primer (Met-Ala-Ser-Met-Thr-Gly-Gly-Gln-Gln-Met-Gly-Arg-Ile-Asn-Ser) (SEQ ID NO:24).

In these experiments, the goal was to generate enzymes that lacked the ability to synthesize DNA, but retained the ability to cleave nucleic acids with a 5' nuclease activity. The act of primed, templated synthesis of DNA is actually a coordinated series of events, so it is possible to disable DNA synthesis by disrupting one event while not affecting the others. These steps include, but are not limited to, primer recognition and binding, dNTP binding and catalysis of the inter-nucleotide phosphodiester bond. Some of the amino acids in the polymerization domain of DNAPEcI have been linked to these functions, but the precise mechanisms are as yet poorly defined.

One way of destroying the polymerizing ability of a DNA polymerase is to delete all or part of the gene segment that encodes that domain for the protein, or to otherwise render the gene incapable of making a complete polymerization domain. Individual mutant enzymes may differ from each other in stability and solubility both inside and outside cells. For instance, in contrast to the 5' nuclease domain of DNAPEcI, which can be released in an active form from the polymerization domain by gentle proteolysis (Setlow and Kornberg, J.

Biol. Chem., 247:232 [1972]), the *Thermus* nuclease domain, when treated similarly, becomes less soluble and the cleavage activity is often lost.

Using the mutant gene shown in Fig. 3B as starting material, several deletion constructs were created. All cloning technologies were standard (Sambrook *et al.*, *supra*) and are summarized briefly, as follows:

Fig. 3C: The mut*Taq* construct was digested with *Pst*I, which cuts once within the polymerase coding region, as indicated, and cuts immediately downstream of the gene in the multiple cloning site of the vector. After release of the fragment between these two sites, the vector was re-ligated, creating an 894-nucleotide deletion, and bringing into frame a stop codon 40 nucleotides downstream of the junction. The nucleotide sequence of this 5' nuclease (clone 4C) is given in SEQ ID NO:9.

Fig. 3D: The mut*Taq* construct was digested with *Nhe*I, which cuts once in the gene at position 2047. The resulting four-nucleotide 5' overhanging ends were filled in, as described above, and the blunt ends were re-ligated. The resulting four-nucleotide insertion changes the reading frame and causes termination of translation ten amino acids downstream of the mutation. The nucleotide sequence of this 5' nuclease (clone 3D) is given in SEQ ID NO:10.

Fig. 3E: The entire mut*Taq* gene was cut from pTTQ18 using *Eco*RI and *Sal*I and cloned into pET-3c, as described above. This clone was digested with *Bst*XI and *Xcm*I, at unique sites that are situated as shown in Fig. 3E. The DNA was treated with the Klenow fragment of DNAPEc1 and dNTPs, which resulted in the 3' overhangs of both sites being trimmed to blunt ends. These blunt ends were ligated together, resulting in an out-of-frame deletion of 1540 nucleotides. An in-frame termination codon occurs 18 triplets past the junction site. The nucleotide sequence of this 5' nuclease (clone 3E) is given in SEQ ID NO:11, with the appropriate leader sequence given in SEQ ID NO:25. It is also referred to as Cleavase® BX.

Fig. 3F: The entire mut*Taq* gene was cut from pTTQ18 using *Eco*RI and *Sal*I and cloned into pET-3c, as described above. This clone was digested with *Bst*XI and *Bam*HI, at unique sites that are situated as shown in the diagram. The DNA was treated with the Klenow fragment of DNAPEc1 and dNTPs, which resulted in the 3' overhang of the *Bst*XI site being trimmed to a blunt end, while the 5' overhang of the *Bam*HI site was filled in to make a blunt end. These ends were ligated together, resulting in an in-frame deletion of 903

nucleotides. The nucleotide sequence of the 5' nuclease (clone 3F) is given in SEQ ID NO:12. It is also referred to as Cleavase® BB.

Fig. 3G: This polymerase is a variant of that shown in Fig. 4E. It was cloned in the plasmid vector pET-21 (Novagen). The non-bacterial promoter from bacteriophage T7, found in this vector, initiates transcription only by T7 RNA polymerase. *See* Studier and Moffatt, *supra*. In a suitable strain, such as (DES)pLYS, the gene for this RNA polymerase is carried on the bacterial genome under control of the *lac* operator. This arrangement has the advantage that expression of the multiple copy gene (on the plasmid) is completely dependent on the expression of T7 RNA polymerase, which is easily suppressed because it is present in a single copy. Because the expression of these mutant genes is under this tightly controlled promoter, potential problems of toxicity of the expressed proteins to the host cells are less of a concern.

The pET-21 vector also features a "His*Tag", a stretch of six consecutive histidine residues that are added on the carboxy terminus of the expressed proteins. The resulting proteins can then be purified in a single step by metal chelation chromatography, using a commercially available (Novagen) column resin with immobilized Ni^{++} ions. The 2.5 ml columns are reusable, and can bind up to 20 mg of the target protein under native or denaturing (guanidine*HCl or urea) conditions.

E. coli (DES)pLYS cells are transformed with the constructs described above using standard transformation techniques, and used to inoculate a standard growth medium (*e.g.*, Luria-Bertani broth). Production of T7 RNA polymerase is induced during log phase growth by addition of IPTG and incubated for a further 12 to 17 hours. Aliquots of culture are removed both before and after induction and the proteins are examined by SDS-PAGE. Staining with Coomassie Blue allows visualization of the foreign proteins if they account for about 3-5% of the cellular protein and do not co-migrate with any of the major protein bands. Proteins that co-migrate with major host protein must be expressed as more than 10% of the total protein to be seen at this stage of analysis.

Some mutant proteins are sequestered by the cells into inclusion bodies. These are granules that form in the cytoplasm when bacteria are made to express high levels of a foreign protein, and they can be purified from a crude lysate, and analyzed by SDS-PAGE to determine their protein content. If the cloned protein is found in the inclusion bodies, it must be released to assay the cleavage and polymerase activities. Different methods of solubilization may be appropriate for different proteins, and a variety of methods are known

(*See e.g.*, Builder & Ogez, U.S. Patent No. 4,511,502 (1985); Olson, U.S. Patent No. 4,518,526 (1985); Olson & Pai, U.S. Patent No. 4,511,503 (1985); and Jones *et al.*, U.S. Patent No. 4,512,922 (1985), all of which are hereby incorporated by reference).

The solubilized protein is then purified on the Ni^{++} column as described above, following the manufacturers instructions (Novagen). The washed proteins are eluted from the column by a combination of imidazole competitor (1 M) and high salt (0.5 M NaCl), and dialyzed to exchange the buffer and to allow denature proteins to refold. Typical recoveries result in approximately 20 μg of specific protein per ml of starting culture. The DNAP mutant is referred to as the Cleavase® BN nuclease and the sequence is given in SEQ ID NO:26 (the amino acid sequence of the Cleavase® BN nuclease is obtained by translating the DNA sequence of SEQ ID NO:26).

2. Modified DNAPTfl Gene

The DNA polymerase gene of *Thermus flavus* was isolated from *the "T. flavus"* AT-62 strain obtained from the American Type Tissue Collection (ATCC 33923). This strain has a different restriction map then does the *T. flavus* strain used to generate the sequence published by Akhmetzjanov and Vakhitov, *supra*. The published sequence is listed as SEQ ID NO:2. No sequence data has been published for the DNA polymerase gene from the AT-62 strain of *T. flavus*.

Genomic DNA from *T. flavus* was amplified using the same primers used to amplify the *T. aquaticus* DNA polymerase gene (SEQ ID NOS:13-14). The approximately 2500 base pair PCR fragment was digested with *Eco*RI and *Bam*HI. The over-hanging ends were made blunt with the Klenow fragment of DNAPEc1 and dNTPs. The resulting approximately 1800 base pair fragment containing the coding region for the N-terminus was ligated into pET-3c, as described above. This construct, clone 4B, is depicted in Fig. 4B. The wild type *T. flavus* DNA polymerase gene is depicted in Fig. 4A. The 4B clone has the same leader amino acids as do the DNAP*Taq* clones 4E and F which were cloned into pET-3c; it is not known precisely where translation termination occurs, but the vector has a strong transcription termination signal immediately downstream of the cloning site.

B. Growth And Induction Of Transformed Cells

Bacterial cells were transformed with the constructs described above using standard transformation techniques and used to inoculate 2 mls of a standard growth medium (*e.g.*,

Luria-Bertani broth). The resulting cultures were incubated as appropriate for the particular strain used, and induced if required for a particular expression system. For all of the constructs depicted in Figs. 3 and 4, the cultures were grown to an optical density (at 600nm wavelength) of 0.5 OD.

To induce expression of the cloned genes, the cultures were brought to a final concentration of 0.4 mM IPTG and the incubations were continued for 12 to 17 hours. Then, 50 μl aliquots of each culture were removed both before and after induction and were combined with 20 μl of a standard gel loading buffer for sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Subsequent staining with Coomassie Blue (Sambrook *et al.*, *supra*) allows visualization of the foreign proteins if they account for about 3-5% of the cellular protein and do not co-migrate with any of the major *E. coli* protein bands. Proteins that do co-migrate with a major host protein must be expressed as more than 10% of the total protein to be seen at this stage of analysis.

C. Heat Lysis And Fractionation

Expressed thermostable proteins (*i.e.*, the 5' nucleases), were isolated by heating crude bacterial cell extracts to cause denaturation and precipitation of the less stable *E. coli* proteins. The precipitated *E. coli* proteins were then, along with other cell debris, removed by centrifugation. Then, 1.7 mls of the culture were pelleted by microcentrifugation at 12,000 to 14,000 rpm for 30 to 60 seconds. After removal of the supernatant, the cells were resuspended in 400 μl of buffer A (50 mM Tris-HC1, pH 7.9, 50 mM dextrose, 1 mM EDTA), re-centrifuged, then resuspended in 80 μl of buffer A with 4 mg/ml lysozyme. The cells were incubated at room temperature for 15 minutes, then combined with 80 μl of buffer B (10 mM Tris-HC1, pH 7.9, 50 mM KCl, 1 mM EDTA, 1 mM PMSF, 0.5% Tween-20, 0.5% Nonidet-P40).

This mixture was incubated at 75°C for 1 hour to denature and precipitate the host proteins. This cell extract was centrifuged at 14,000 rpm for 15 minutes at 4°C, and the supernatant was transferred to a fresh tube. An aliquot of 0.5 to 1 μl of this supernatant was used directly in each test reaction, and the protein content of the extract was determined by subjecting 7 μl to electrophoretic analysis, as above. The native recombinant *Taq* DNA polymerase (Engelke, Anal. Biochem., 191:396 [1990]), and the double point mutation protein shown in Fig. 3B are both soluble and active at this point.

The foreign protein may not be detected after the heat treatments due to sequestration of the foreign protein by the cells into inclusion bodies. These are granules that form in the cytoplasm when bacteria are made to express high levels of a foreign protein, and they can be purified from a crude lysate, and analyzed SDS PAGE to determine their protein content. Many methods have been described in the literature, and one approach is described below.

D. Isolation And Solubilization Of Inclusion Bodies

A small culture was grown and induced as described above. A 1.7 ml aliquot was pelleted by brief centrifugation, and the bacterial cells were resuspended in 100 µl of Lysis buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 100 mM NaCl). Then, 2.5 µl of 20 mM PMSF were added for a final concentration of 0.5 mM, and lysozyme was added to a concentration of 1.0 mg/ml. The cells were incubated at room temperature for 20 minutes, deoxycholic acid was added to 1mg/ml (1 µl of 100 mg/ml solution), and the mixture was further incubated at 37°C for about 15 minutes or until viscous. DNAse I was added to 10 µg/ml and the mixture was incubated at room temperature for about 30 minutes or until it was no longer viscous.

From this mixture the inclusion bodies were collected by centrifugation at 14,000 rpm for 15 minutes at 4°C, and the supernatant was discarded. The pellet was resuspended in 100 µl of lysis buffer with 10 mM EDTA (pH 8.0) and 0.5% Triton X-100. After 5 minutes at room temperature, the inclusion bodies were pelleted as before, and the supernatant was saved for later analysis. The inclusion bodies were resuspended in 50 µl of distilled water, and 5 µl was combined with SDS gel loading buffer (which dissolves the inclusion bodies) and analyzed electrophoretically, along with an aliquot of the supernatant.

If the cloned protein is found in the inclusion bodies, it may be released to assay the cleavage and polymerase activities and the method of solubilization must be compatible with the particular activity. Different methods of solubilization may be appropriate for different proteins, and a variety of methods are discussed in *Molecular Cloning* (Sambrook *et al.*, *supra*). The following is an adaptation used for several of the isolates used in the development of the present invention.

Twenty µl of the inclusion body-water suspension were pelleted by centrifugation at 14,000 rpm for 4 minutes at room temperature, and the supernatant was discarded. To further wash the inclusion bodies, the pellet was resuspended in 20 µl of lysis buffer with 2M urea, and incubated at room temperature for one hour. The washed inclusion bodies were then

resuspended in 2 µl of lysis buffer with 8 M urea; the solution clarified visibly as the inclusion bodies dissolved. Undissolved debris was removed by centrifugation at 14,000 rpm for 4 minutes at room temperature, and the extract supernatant was transferred to a fresh tube.

To reduce the urea concentration, the extract was diluted into KH_2PO_4. A fresh tube was prepared containing 180 µl of 50 mM KH_2PO_4, pH 9.5, 1 mM EDTA and 50 mM NaCl. A 2 µl aliquot of the extract was added and vortexed briefly to mix. This step was repeated until all of the extract had been added for a total of 10 additions. The mixture was allowed to sit at room temperature for 15 minutes, during which time some precipitate often forms. Precipitates were removed by centrifugation at 14,000 rpm, for 15 minutes at room temperature, and the supernatant was transferred to a fresh tube. To the 200 µl of protein in the KH_2PO_4 solution, 140-200 µl of saturated $(NH_4)_2SO_4$ were added, so that the resulting mixture was about 41% to 50% saturated $(NH_4)_2SO_4$. The mixture was chilled on ice for 30 minutes to allow the protein to precipitate, and the protein was then collected by centrifugation at 14,000 rpm, for 4 minutes at room temperature. The supernatant was discarded, and the pellet was dissolved in 20 µl Buffer C (20 mM HEPES, pH 7.9, 1 mM EDTA, 0.5% PMSF, 25 mM KCl and 0.5 % each of Tween-20 and Nonidet P 40). The protein solution was centrifuged again for 4 minutes to pellet insoluble materials, and the supernatant was removed to a fresh tube. The protein contents of extracts prepared in this manner were visualized by resolving 1-4 µl by SDS-PAGE; 0.5 to 1 µl of extract was tested in the cleavage and polymerization assays as described.

E. Protein Analysis For Presence Of Nuclease And Synthetic Activity

The 5' nucleases described above and shown in Figs. 3 and 4 were analyzed by the following methods.

1. Structure Specific Nuclease Assay

A candidate modified polymerase is tested for 5' nuclease activity by examining its ability to catalyze structure-specific cleavages. By the term "cleavage structure" as used herein, is meant a nucleic acid structure which is a substrate for cleavage by the 5' nuclease activity of a DNAP.

The polymerase is exposed to test complexes that have the structures shown in Fig. 15. Testing for 5' nuclease activity involves three reactions: 1) a primer-directed cleavage (Fig. 15B) is performed because it is relatively insensitive to variations in the salt

concentration of the reaction and can, therefore, be performed in whatever solute conditions the modified enzyme requires for activity; this is generally the same conditions preferred by unmodified polymerases; 2) a similar primer-directed cleavage is performed in a buffer which permits primer-independent cleavage (*i.e.*, a low salt buffer), to demonstrate that the enzyme is viable under these conditions; and 3) a primer-independent cleavage (Fig. 15A) is performed in the same low salt buffer.

The bifurcated duplex is formed between a substrate strand and a template strand as shown in Fig. 15. By the term "substrate strand" as used herein, is meant that strand of nucleic acid in which the cleavage mediated by the 5' nuclease activity occurs. The substrate strand is always depicted as the top strand in the bifurcated complex which serves as a substrate for 5' nuclease cleavage (Fig. 15). By the term "template strand" as used herein, is meant the strand of nucleic acid which is at least partially complementary to the substrate strand and which anneals to the substrate strand to form the cleavage structure. The template strand is always depicted as the bottom strand of the bifurcated cleavage structure (Fig. 15). If a primer (a short oligonucleotide of 19 to 30 nucleotides in length) is added to the complex, as when primer-dependent cleavage is to be tested, it is designed to anneal to the 3' arm of the template strand (Fig. 15B). Such a primer would be extended along the template strand if the polymerase used in the reaction has synthetic activity.

The cleavage structure may be made as a single hairpin molecule, with the 3' end of the target and the 5' end of the pilot joined as a loop as shown in Fig. 15E. A primer oligonucleotide complementary to the 3' arm is also required for these tests so that the enzyme's sensitivity to the presence of a primer may be tested.

Nucleic acids to be used to form test cleavage structures can be chemically synthesized, or can be generated by standard recombinant DNA techniques. By the latter method, the hairpin portion of the molecule can be created by inserting into a cloning vector duplicate copies of a short DNA segment, adjacent to each other but in opposing orientation. The double-stranded fragment encompassing this inverted repeat, and including enough flanking sequence to give short (about 20 nucleotides) unpaired 5' and 3' arms, can then be released from the vector by restriction enzyme digestion, or by PCR performed with an enzyme lacking a 5' exonuclease (*e.g.*, the Stoffel fragment of Amplitaq™ DNA polymerase, Vent™ DNA polymerase).

The test DNA can be labeled on either end, or internally, with either a radioisotope, or with a non-isotopic tag. Whether the hairpin DNA is a synthetic single strand or a cloned

double strand, the DNA is heated prior to use to melt all duplexes. When cooled on ice, the structure depicted in Fig. 16E is formed, and is stable for sufficient time to perform these assays.

To test for primer-directed cleavage (Reaction 1), a detectable quantity of the test molecule (typically 1-100 fmol of ^{32}P-labeled hairpin molecule) and a 10 to 100-fold molar excess of primer are placed in a buffer known to be compatible with the test enzyme. For Reaction 2, where primer-directed cleavage is performed under condition which allow primer-independent cleavage, the same quantities of molecules are placed in a solution that is the same as the buffer used in Reaction 1 regarding pH, enzyme stabilizers (*e.g.*, bovine serum albumin, nonionic detergents, gelatin) and reducing agents (*e.g.*, dithiothreitol, 2-mercaptoethanol) but that replaces any monovalent cation salt with 20 mM KCl; 20 mM KCl is the demonstrated optimum for primer-independent cleavage. Buffers for enzymes, such as DNAPEc1, that usually operate in the absence of salt are not supplemented to achieve this concentration. To test for primer-independent cleavage (Reaction 3) the same quantity of the test molecule, but no primer, are combined under the same buffer conditions used for Reaction 2.

All three test reactions are then exposed to enough of the enzyme that the molar ratio of enzyme to test complex is approximately 1:1. The reactions are incubated at a range of temperatures up to, but not exceeding, the temperature allowed by either the enzyme stability or the complex stability, whichever is lower, up to 80°C for enzymes from thermophiles, for a time sufficient to allow cleavage (10 to 60 minutes). The products of Reactions 1, 2 and 3 are resolved by denaturing polyacrylamide gel electrophoresis, and visualized by autoradiography or by a comparable method appropriate to the labeling system used. Additional labeling systems include chemiluminescence detection, silver or other stains, blotting and probing and the like. The presence of cleavage products is indicated by the presence of molecules which migrate at a lower molecular weight than does the uncleaved test structure. These cleavage products indicate that the candidate polymerase has structure-specific 5' nuclease activity.

To determine whether a modified DNA polymerase has substantially the same 5' nuclease activity as that of the native DNA polymerase, the results of the above-described tests are compared with the results obtained from these tests performed with the native DNA polymerase. By "substantially the same 5' nuclease activity" it is meant that the modified polymerase and the native polymerase will both cleave test molecules in the same manner. It

is not necessary that the modified polymerase cleave at the same rate as the native DNA polymerase.

Some enzymes or enzyme preparations may have other associated or contaminating activities that may be functional under the cleavage conditions described above and that may interfere with 5' nuclease detection. Reaction conditions can be modified in consideration of these other activities, to avoid destruction of the substrate, or other masking of the 5' nuclease cleavage and its products. For example, the DNA polymerase I of *E. coli* (Pol I), in addition to its polymerase and 5' nuclease activities, has a 3' exonuclease that can degrade DNA in a 3' to 5' direction. Consequently, when the molecule in Fig. 15E is exposed to this polymerase under the conditions described above, the 3' exonuclease quickly removes the unpaired 3' arm, destroying the bifurcated structure required of a substrate for the 5' exonuclease cleavage and no cleavage is detected. The true ability of Pol I to cleave the structure can be revealed if the 3' exonuclease is inhibited by a change of conditions (*e.g.*, pH), mutation, or by addition of a competitor for the activity. Addition of 500 pmoles of a single-stranded competitor oligonucleotide, unrelated to the Fig. 15E structure, to the cleavage reaction with Pol I effectively inhibits the digestion of the 3' arm of the Fig. 15E structure without interfering with the 5' exonuclease release of the 5' arm. The concentration of the competitor is not critical, but should be high enough to occupy the 3' exonuclease for the duration of the reaction.

Similar destruction of the test molecule may be caused by contaminants in the candidate polymerase preparation. Several sets of the structure specific nuclease reactions may be performed to determine the purity of the candidate nuclease and to find the window between under and over exposure of the test molecule to the polymerase preparation being investigated.

The above described modified polymerases were tested for 5' nuclease activity as follows: Reaction 1 was performed in a buffer of 10 mM Tris-Cl, pH 8.5 at 20°C, 1.5 mM $MgCl_2$ and 50 mM KCl and in Reaction 2 the KCl concentration was reduced to 20 mM. In Reactions 1 and 2, 10 fmoles of the test substrate molecule shown in Fig. 15E were combined with 1 pmole of the indicated primer and 0.5 to 1.0 μl of extract containing the modified polymerase (prepared as described above). This mixture was then incubated for 10 minutes at 55°C. For all of the mutant polymerases tested these conditions were sufficient to give complete cleavage. When the molecule shown in Fig. 15E was labeled at the 5' end, the released 5' fragment, 25 nucleotides long, was conveniently resolved on a 20%

polyacrylamide gel (19:1 cross-linked) with 7 M urea in a buffer containing 45 mM Tris-borate pH 8.3, 1.4 mM EDTA. Clones 3C-F and 4B exhibited structure-specific cleavage comparable to that of the unmodified DNA polymerase. Additionally, clones 3E, 3F and 3G have the added ability to cleave DNA in the absence of a 3' arm as discussed above. Representative cleavage reactions are shown in Fig. 16.

For the reactions shown in Fig. 16, the mutant polymerase clones 3E (*Taq* mutant) and 4B (*Tfl* mutant) were examined for their ability to cleave the hairpin substrate molecule shown in Fig. 15E. The substrate molecule was labeled at the 5' terminus with ^{32}P. Ten fmoles of heat-denatured, end-labeled substrate DNA and 0.5 units of DNAP*Taq* (lane 1) or 0.5 μl of 3E or 4B extract (Fig. 16, lanes 2-7, extract was prepared as described above) were mixed together in a buffer containing 10 mM Tris-Cl, pH 8.5, 50 mM KCl and 1.5 mM $MgCl_2$. The final reaction volume was 10 μl. Reactions shown in lanes 4 and 7 contain in addition 50 μM of each dNTP. Reactions shown in lanes 3, 4, 6 and 7 contain 0.2 μM of the primer oligonucleotide (complementary to the 3' arm of the substrate and shown in Fig. 15E). Reactions were incubated at 55° C for 4 minutes. Reactions were stopped by the addition of 8 μl of 95% formamide containing 20 mM EDTA and 0.05% marker dyes per 10 μl reaction volume. Samples were then applied to 12% denaturing acrylamide gels. Following electrophoresis, the gels were autoradiographed. Fig. 16 shows that clones 3E and 4B exhibit cleavage activity similar to that of the native DNAP*Taq*. Note that some cleavage occurs in these reactions in the absence of the primer. When long hairpin structure, such as the one used here (Fig. 15E), are used in cleavage reactions performed in buffers containing 50 mM KCl a low level of primer-independent cleavage is seen. Higher concentrations of KCl suppress, but do not eliminate, this primer-independent cleavage under these conditions.

2. Assay For Synthetic Activity

The ability of the modified enzyme or proteolytic fragments is assayed by adding the modified enzyme to an assay system in which a primer is annealed to a template and DNA synthesis is catalyzed by the added enzyme. Many standard laboratory techniques employ such an assay. For example, nick translation and enzymatic sequencing involve extension of a primer along a DNA template by a polymerase molecule.

In a preferred assay for determining the synthetic activity of a modified enzyme an oligonucleotide primer is annealed to a single-stranded DNA template (*e.g.*, bacteriophage M13 DNA), and the primer/template duplex is incubated in the presence of the modified

polymerase in question, deoxynucleoside triphosphates (dNTPs) and the buffer and salts known to be appropriate for the unmodified or native enzyme. Detection of either primer extension (by denaturing gel electrophoresis) or dNTP incorporation (by acid precipitation or chromatography) is indicative of an active polymerase. A label, either isotopic or non-isotopic, is preferably included on either the primer or as a dNTP to facilitate detection of polymerization products. Synthetic activity is quantified as the amount of free nucleotide incorporated into the growing DNA chain and is expressed as amount incorporated per unit of time under specific reaction conditions.

Representative results of an assay for synthetic activity is shown in Fig. 17. The synthetic activity of the mutant DNAP*Taq* clones 3B-F was tested as follows: A master mixture of the following buffer was made: 1.2X PCR buffer (1X PCR buffer contains 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris-Cl, pH 8.5 and 0.05% each Tween 20 and Nonidet P40), 50 μM each of dGTP, dATP and dTTP, 5 μM dCTP and 0.125 μM α-^{32}P-dCTP at 600 Ci/mmol. Before adjusting this mixture to its final volume, it was divided into two equal aliquots. One received distilled water up to a volume of 50 μl to give the concentrations above. The other received 5 μg of single-stranded M13mp18 DNA (approximately 2.5 pmol or 0.05 μM final concentration) and 250 pmol of M13 sequencing primer (5 μM final concentration) and distilled water to a final volume of 50 μl. Each cocktail was warmed to 75°C for 5 minutes and then cooled to room temperature. This allowed the primers to anneal to the DNA in the DNA-containing mixtures.

For each assay, 4 μl of the cocktail with the DNA was combined with 1 μl of the mutant polymerase, prepared as described, or 1 unit of DNAP*Taq* (Perkin Elmer) in 1 μl of dH_2O. A "no DNA" control was done in the presence of the DNAP*Taq* (Fig. 17, lane 1), and a "no enzyme" control was done using water in place of the enzyme (lane 2). Each reaction was mixed, then incubated at room temperature (approx. 22°C) for 5 minutes, then at 55°C for 2 minutes, then at 72°C for 2 minutes. This step incubation was done to detect polymerization in any mutants that might have optimal temperatures lower than 72°C. After the final incubation, the tubes were spun briefly to collect any condensation and were placed on ice. One μl of each reaction was spotted at an origin 1.5 cm from the bottom edge of a polyethyleneimine (PEI) cellulose thin layer chromatography plate and allowed to dry. The chromatography plate was run in 0.75 M NaH_2PO_4, pH 3.5, until the buffer front had run approximately 9 cm from the origin. The plate was dried, wrapped in plastic wrap, marked with luminescent ink, and exposed to X-ray film. Incorporation was detected as counts that

stuck where originally spotted, while the unincorporated nucleotides were carried by the salt solution from the origin.

Comparison of the locations of the counts with the two control lanes confirmed the lack of polymerization activity in the mutant preparations. Among the modified DNAP*Taq* clones, only clone 3B retains any residual synthetic activity as shown in Fig. 17.

EXAMPLE 3

5' Nucleases Derived From Thermostable DNA Polymerases Can Cleave Short Hairpin Structures With Specificity

The ability of the 5' nucleases to cleave hairpin structures to generate a cleaved hairpin structure suitable as a detection molecule was examined. The structure and sequence of the hairpin test molecule is shown in Fig. 18A (SEQ ID NO:15). The oligonucleotide (labeled "primer" in Fig. 18A, SEQ ID NO:22) is shown annealed to its complementary sequence on the 3' arm of the hairpin test molecule. The hairpin test molecule was single-end labeled with ^{32}P using a labeled T7 promoter primer in a polymerase chain reaction. The label is present on the 5' arm of the hairpin test molecule and is represented by the star in Fig. 18A.

The cleavage reaction was performed by adding 10 fmoles of heat-denatured, end-labeled hairpin test molecule, 0.2 μM of the primer oligonucleotide (complementary to the 3' arm of the hairpin), 50 μM of each dNTP and 0.5 units of DNAP*Taq* (Perkin Elmer) or 0.5 μl of extract containing a 5' nuclease (prepared as described above) in a total volume of 10 μl in a buffer containing 10 mM Tris-Cl, pH 8.5, 50 mM KCl and 1.5 mM $MgCl_2$. Reactions shown in lanes 3, 5 and 7 were run in the absence of dNTPs.

Reactions were incubated at 55° C for 4 minutes. Reactions were stopped at 55° C by the addition of 8 μl of 95% formamide with 20 mM EDTA and 0.05% marker dyes per 10 μl reaction volume. Samples were not heated before loading onto denaturing polyacrylamide gels (10% polyacrylamide, 19:1 crosslinking, 7 M urea, 89 mM Tris-borate, pH 8.3, 2.8 mM EDTA). The samples were not heated to allow for the resolution of single-stranded and re-duplexed uncleaved hairpin molecules.

Fig. 18B shows that altered polymerases lacking any detectable synthetic activity cleave a hairpin structure when an oligonucleotide is annealed to the single-stranded 3' arm of the hairpin to yield a single species of cleaved product (Fig. 18B, lanes 3 and 4). 5' nucleases, such as clone 3D, shown in lanes 3 and 4, produce a single cleaved product even

in the presence of dNTPs. 5' nucleases which retain a residual amount of synthetic activity (less than 1% of wild type activity) produce multiple cleavage products as the polymerase can extend the oligonucleotide annealed to the 3' arm of the hairpin thereby moving the site of cleavage (clone 3B, lanes 5 and 6). Native DNA*Taq* produces even more species of cleavage products than do mutant polymerases retaining residual synthetic activity and additionally converts the hairpin structure to a double-stranded form in the presence of dNTPs due to the high level of synthetic activity in the native polymerase (Fig. 18B, lane 8).

EXAMPLE 4

Cleavage Of Linear Nucleic Acid Substrates

From the above, it should be clear that native (*i.e.*, "wild type") thermostable DNA polymerases are capable of cleaving hairpin structures in a specific manner and that this discovery can be applied with success to a detection assay. In this example, the mutant DNAPs of the present invention are tested against three different cleavage structures shown in Fig. 20A. Structure 1 in Fig. 20A is simply single stranded 206-mer (the preparation and sequence information for which was discussed in Example 1C). Structures 2 and 3 are duplexes; structure 2 is the same hairpin structure as shown in Fig. 11A (bottom), while structure 3 has the hairpin portion of structure 2 removed.

The cleavage reactions comprised 0.01 pmoles of the resulting substrate DNA, and 1 pmole of pilot oligonucleotide in a total volume of 10 μl of 10 mM Tris-Cl, pH 8.3, 100 mM KCl, 1 mM $MgCl_2$. Reactions were incubated for 30 minutes at 55°C, and stopped by the addition of 8 μl of 95% formamide with 20 mM EDTA and 0.05% marker dyes. Samples were heated to 75°C for 2 minutes immediately before electrophoresis through a 10% polyacrylamide gel (19:1 cross link), with 7M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA.

The results were visualized by autoradiography and are shown in Fig. 20B with the enzymes indicated as follows: I is native *Taq* DNAP; II is native *Tfl* DNAP; III is Cleavase® BX shown in Fig. 3E; IV is Cleavase® BB shown in Fig. 3F; V is the mutant shown in Fig. 4B; and VI is Cleavase® BN shown in Fig. 3G.

Structure 2 was used to "normalize" the comparison. For example, it was found that it took 50 ng of *Taq* DNAP and 300 ng of Cleavase® BN to give similar amounts of cleavage of Structure 2 in thirty (30) minutes. Under these conditions native *Taq* DNAP is unable to

cleave Structure 3 to any significant degree. Native *Tfl* DNAP cleaves Structure 3 in a manner that creates multiple products.

By contrast, all of the mutants tested cleave the linear duplex of Structure 3. This finding indicates that this characteristic of the mutant DNA polymerases is consistent of thermostable polymerases across thermophilic species.

EXAMPLE 5

5' Exonucleolytic Cleavage ("Nibbling") By Thermostable DNAPs

It has been found that thermostable DNAPs, including those of the present invention, have a true 5' exonuclease capable of nibbling the 5' end of a linear duplex nucleic acid structures. In this Example, the 206 base pair DNA duplex substrate is again employed (*See*, Example 1C). In this case, it was produced by the use of one ^{32}P-labeled primer and one unlabeled primer in a polymerase chain reaction. The cleavage reactions comprised 0.01 pmoles of heat-denatured, end-labeled substrate DNA (with the unlabeled strand also present), 5 pmoles of pilot oligonucleotide (see pilot oligos in Fig. 11A) and 0.5 units of DNAPTaq or 0.5 μ of Cleavase® BB in the *E. coli* extract (see above), in a total volume of 10 μl of 10 mM Tris•Cl, pH 8.5, 50 mM KCl, 1.5 mM $MgCl_2$.

Reactions were initiated at 65°C by the addition of pre-warmed enzyme, then shifted to the final incubation temperature for 30 minutes. The results are shown in Fig. 21A. Samples in lanes 1-4 are the results with native *Taq* DNAP, while lanes 5-8 shown the results with Cleavase® BB. The reactions for lanes 1, 2, 5, and 6 were performed at 65°C and reactions for lanes 3, 4, 7, and 8 were performed at 50°C and all were stopped at temperature by the addition of 8 μl of 95% formamide with 20 mM EDTA and 0.05% marker dyes. Samples were heated to 75°C for 2 minutes immediately before electrophoresis through a 10% acrylamide gel (19:1 cross-linked), with 7 M urea, in a buffer of 45 mM Tris•Borate, pH 8.3, 1.4 mM EDTA. The expected product in reactions 1, 2, 5, and 6 is 85 nucleotides long; in reactions 3 and 7, the expected product is 27 nucleotides long. Reactions 4 and 8 were performed without pilot, and should remain at 206 nucleotides. The faint band seen at 24 nucleotides is residual end-labeled primer from the PCR.

The surprising result is that Cleavase® BB under these conditions causes all of the label to appear in a very small species, suggesting the possibility that the enzyme completely hydrolyzed the substrate. To determine the composition of the fastest-migrating band seen in lanes 5-8 (reactions performed with the deletion mutant), samples of the 206 base pair duplex

were treated with either T7 gene 6 exonuclease (USB) or with calf intestine alkaline phosphatase (Promega), according to manufacturers' instructions, to produce either labeled mononucleotide (lane a of Fig. 21B) or free ^{32}P-labeled inorganic phosphate (lane b of Fig. 21B), respectively. These products, along with the products seen in lane 7 of panel A were resolved by brief electrophoresis through a 20% acrylamide gel (19:1 cross-link), with 7 M urea, in a buffer of 45 mM Tris•Borate, pH 8.3, 1.4 mM EDTA. Cleavase® BB is thus capable of converting the substrate to mononucleotides.

EXAMPLE 6

Nibbling Is Duplex Dependent

The nibbling by Cleavase® BB is duplex dependent. In this Example, internally labeled, single strands of the 206-mer were produced by 15 cycles of primer extension incorporating α-^{32}P labeled dCTP combined with all four unlabeled dNTPs, using an unlabeled 206-bp fragment as a template. Single and double stranded products were resolved by electrophoresis through a non-denaturing 6% polyacrylamide gel (29:1 cross-link) in a buffer of 45 mM Tris•Borate, pH 8.3, 1.4 mM EDTA, visualized by autoradiography, excised from the gel, eluted by passive diffusion, and concentrated by ethanol precipitation.

The cleavage reactions comprised 0.04 pmoles of substrate DNA, and 2 μl of Cleavase® BB (in an *E. coli* extract as described above) in a total volume of 40 μl of 10 mM Tris•Cl, pH 8.5, 50 mM KCl, 1.5 mM $MgCl_2$. Reactions were initiated by the addition of pre-warmed enzyme; 10 μl aliquots were removed at 5, 10, 20, and 30 minutes, and transferred to prepared tubes containing 8 μl of 95% formamide with 30 mM EDTA and 0.05% marker dyes. Samples were heated to 75°C for 2 minutes immediately before electrophoresis through a 10% acrylamide gel (19:1 cross-linked), with 7 M urea, in a buffer of 45 mM Tris•Borate, pH 8.3, 1.4 mM EDTA. Results were visualized by autoradiography as shown in Fig. 22. Clearly, the cleavage by Cleavase® BB depends on a duplex structure; no cleavage of the single strand structure is detected whereas cleavage of the 206-mer duplex is complete.

EXAMPLE 7

Nibbling Can Be Target Directed

The nibbling activity of the DNAPs of the present invention can be employed with success in a detection assay. One embodiment of such an assay is shown in Fig. 23. In this assay, a labelled oligo is employed that is specific for a target sequence. The oligo is in excess of the target so that hybridization is rapid. In this embodiment, the oligo contains two fluorescein labels whose proximity on the oligo causes their emission to be quenched. When the DNAP is permitted to nibble the oligo the labels separate and are detectable. The shortened duplex is destabilized and disassociates. Importantly, the target is now free to react with an intact labelled oligo. The reaction can continue until the desired level of detection is achieved. An analogous, although different, type of cycling assay has been described employing lambda exonuclease. *See* C.G. Copley and C. Boot, *BioTechniques* 13:888 (1992).

The success of such an assay depends on specificity. In other words, the oligo must hybridize to the specific target. It is also preferred that the assay be sensitive; the oligo ideally should be able to detect small amounts of target. Fig. 24A shows a 5'-end ^{32}P-labelled primer bound to a plasmid target sequence. In this case, the plasmid was pUC19 (commercially available) which was heat denatured by boiling two (2) minutes and then quick chilling. The primer is a 21-mer (SEQ ID NO:28). The enzyme employed was Cleavase® BX (a dilution equivalent to 5 x 10^{-3} µl extract) in 100 mM KCl, 10 mM Tris-Cl, pH 8.3, 2 mM $MnCl_2$. The reaction was performed at 55°C for sixteen (16) hours with or without genomic background DNA (from chicken blood). The reaction was stopped by the addition of 8 µl of 95% formamide with 20 mM EDTA and marker dyes.

The products of the reaction were resolved by PAGE (10% polyacrylamide, 19:1 cross link, 1X TBE) as seen in Fig. 24B. Lane "M" contains the labelled 21-mer. Lanes 1-3 contain no specific target, although Lanes 2 and 3 contain 100 ng and 200 ng of genomic DNA, respectively. Lanes 4, 5 and 6 all contain specific target with either 0 ng, 100 ng, or 200 ng of genomic DNA, respectively. It is clear that conversion to mononucleotides occurs in Lanes 4, 5 and 6 regardless of the presence or amount of background DNA. Thus, the nibbling can be target directed and specific.

EXAMPLE 8

Cleavase Purification

As noted above, expressed thermostable proteins (*i.e.*, the 5' nucleases), were isolated by crude bacterial cell extracts. The precipitated *E. coli* proteins were then, along with other cell debris, removed by centrifugation. In this Example, cells expressing the BN clone were cultured and collected (500 grams). For each gram (wet weight) of *E. coli*, 3 ml of lysis buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 100μM NaCl) was added. The cells were lysed with 200 μg/ml lysozyme at room temperature for 20 minutes. Thereafter deoxycholic acid was added to make a 0.2% final concentration and the mixture was incubated 15 minutes at room temperature.

The lysate was sonicated for approximately 6-8 minutes at 0°C. The precipitate was removed by centrifugation (39,000*g* for 20 minutes). Polyethyleneimine was added (0.5%) to the supernatant and the mixture was incubated on ice for 15 minutes.
The mixture was centrifuged (5,000*g* for 15 minutes) and the supernatant was retained. This was heated for 30 minutes at 60°C and then centrifuged again (5,000*g* for 15 minutes) and the supernatant was again retained.

The supernatant was precipitated with 35% ammonium sulfate at 4°C for 15 minutes. The mixture was then centrifuged (5,000*g* for 15 minutes) and the supernatant was removed. The precipitate was then dissolved in 0.25M KCl, 20 Tris pH 7.6, 0.2% Tween and 0.1 EDTA) and then dialyzed against Binding Buffer (8X Binding Buffer comprises: 40mM imidazole, 4M NaCl, 160mM Tris-HCl, pH 7.9).

The solubilized protein is then purified on the Ni^{++} column (Novagen). The Binding Buffer is allows to drain to the top of the column bed and load the column with the prepared extract. A flow rate of about 10 column volumes per hour is optimal for efficient purification. If the flow rate is too fast, more impurities will contaminate the eluted fraction.

The column is washed with 25 ml (10 volumes) of 1X Binding Buffer and then washed with 15 ml (6 volumes) of 1X Wash Buffer (8X Wash Buffer comprises: 480mM imidazole, 4 M NaCl, 160 mM Tris-HCl, pH 7.9). The bound protein was eluted with 15 ml (6 volumes) of 1X Elute Buffer (4X Elute Buffer comprises: 4mM imidazole, 2 M NaCl, 80 mM Tris-HCl, pH 7.9). Protein is then reprecipitated with 35% ammonium sulfate as above. The precipitate was then dissolved and dialyzed against: 20 mM Tris, 100 mM KCl, 1 mM EDTA). The solution was brought up to 0.1% each of Tween 20 and NP-40 and stored at 4°C.

EXAMPLE 9

The Use Of Various Divalent Cations In The Cleavage Reaction Influences The Nature Of The Resulting Cleavage Products

In comparing the 5' nucleases generated by the modification and/or deletion of the C-terminal polymerization domain of *Thermus aquaticus* DNA polymerase (DNAPTaq), as diagrammed in Fig. 3B-G, significant differences in the strength of the interactions of these proteins with the 3' end of primers located upstream of the cleavage site (as depicted in Fig. 5) were noted. In describing the cleavage of these structures by Pol I-type DNA polymerases (*See,* Example 1, and Lyamichev *et al.*, Science 260:778 [1993]), it was observed that in the absence of a primer, the location of the junction between the double-stranded region and the single-stranded 5' and 3' arms determined the site of cleavage, but in the presence of a primer, the location of the 3' end of the primer became the determining factor for the site of cleavage. It was postulated that this affinity for the 3' end was in accord with the synthesizing function of the DNA polymerase.

Structure 2, shown in Fig. 20A, was used to test the effects of a 3' end proximal to the cleavage site in cleavage reactions comprising several different solutions (*e.g.*, solutions containing different salts [KCl or NaCl], different divalent cations [Mn^{2+} or Mg^{2+}], etc.) as well as the use of different temperatures for the cleavage reaction. When the reaction conditions were such that the binding of the enzyme (*e.g.*, a DNAP comprising a 5' nuclease, a modified DNAP or a 5' nuclease) to the 3' end (of the pilot oligonucleotide) near the cleavage site was strong, the structure shown is cleaved at the site indicated in Fig. 20A. This cleavage releases the unpaired 5' arm and leaves a nick between the remaining portion of the target nucleic acid and the folded 3' end of the pilot oligonucleotide. In contrast, when the reaction conditions are such that the binding of the DNAP (comprising a 5' nuclease) to the 3' end was weak, the initial cleavage was as described above, but after the release of the 5' arm, the remaining duplex is digested by the exonuclease function of the DNAP.

One way of weakening the binding of the DNAP to the 3' end is to remove all or part of the domain to which at least some of this function has been attributed. Some of 5' nucleases created by deletion of the polymerization domain of DNAPTaq have enhanced true exonuclease function, as demonstrated in Example 5.

The affinity of these types of enzymes (*i.e.*, 5' nucleases associated with or derived from DNAPs) for recessed 3' ends may also be affected by the identity of the divalent cation present in the cleavage reaction. It was demonstrated by Longley *et al.* (Nucl. Acids Res.,

18:7317 [1990]) that the use of $MnCl_2$ in a reaction with DNAPTaq enabled the polymerase to remove nucleotides from the 5' end of a primer annealed to a template, albeit inefficiently. Similarly, by examination of the cleavage products generated using Structure 2 from Fig. 20A, as described above, in a reaction containing either DNAPTaq or the Cleavase® BB nuclease, it was observed that the substitution of $MnCl_2$ for $MgCl_2$ in the cleavage reaction resulted in the exonucleolytic "nibbling" of the duplex downstream of the initial cleavage site. While not limiting the invention to any particular mechanism, it is thought that the substitution of $MnCl_2$ for $MgCl_2$ in the cleavage reaction lessens the affinity of these enzymes for recessed 3' ends.

In all cases, the use of $MnCl_2$ enhances the 5' nuclease function, and in the case of the Cleavase® BB nuclease, a 50- to 100-fold stimulation of the 5' nuclease function is seen. Thus, while the exonuclease activity of these enzymes was demonstrated above in the presence of $MgCl_2$, the assays described below show a comparable amount of exonuclease activity using 50 to 100-fold less enzyme when $MnCl_2$ is used in place of $MgCl_2$. When these reduced amounts of enzyme are used in a reaction mixture containing $MgCl_2$, the nibbling or exonuclease activity is much less apparent than that seen in Examples 5-7.

Similar effects are observed in the performance of the nucleic acid detection assay described in Examples 10-39 below when reactions performed in the presence of either $MgCl_2$ or $MnCl_2$ are compared. In the presence of either divalent cation, the presence of the Invader™ oligonucleotide (described below) forces the site of cleavage into the probe duplex, but in the presence of $MnCl_2$ the probe duplex can be further nibbled producing a ladder of products that are visible when a 3' end label is present on the probe oligonucleotide. When the Invader™ oligonucleotide is omitted from a reaction containing Mn^{2+}, the probe is nibbled from the 5' end. Mg^{2+}-based reactions display minimal nibbling of the probe oligonucleotide. In any of these cases, the digestion of the probe is dependent upon the presence of the target nucleic acid. In the examples below, the ladder produced by the enhanced nibbling activity observed in the presence of Mn^{2+} is used as a positive indicator that the probe oligonucleotide has hybridized to the target sequence.

EXAMPLE 10

Invasive 5' Endonucleolytic Cleavage By Thermostable 5' Nucleases In The Absence of Polymerization

As described in the Examples above, 5' nucleases cleave near the junction between single-stranded and base-paired regions in a bifurcated duplex, usually about one base pair into the base-paired region. In this Example, it is shown that thermostable 5' nucleases, including those of the present invention (*e.g.*, Cleavase® BN nuclease, Cleavase® A/G nuclease), have the ability to cleave a greater distance into the base paired region when provided with an upstream oligonucleotide bearing a 3' region that is homologous to a 5' region of the subject duplex, as shown in Fig. 26.

Fig. 26 shows a synthetic oligonucleotide which was designed to fold upon itself which consists of the following sequence: 5'-GTTCTCTGCTCTCTGGTCGCTG TCTCGCTTGTGAAACAAGCGAGACAGCGTGGTCTCTCG-3' (SEQ ID NO:29). This oligonucleotide is referred to as the "S-60 Hairpin." The 15 basepair hairpin formed by this oligonucleotide is further stabilized by a "tri-loop" sequence in the loop end (*i.e.*, three nucleotides form the loop portion of the hairpin) (Hiraro *et al.*, Nucleic Acids Res., 22(4):576 [1994]). Fig. 26 also show the sequence of the P-15 oligonucleotide and the location of the region of complementarity shared by the P-15 and S-60 hairpin oligonucleotides. The sequence of the P-15 oligonucleotide is 5'-CGAGAGACCACGCTG-3' (SEQ ID NO:30). As discussed in detail below, the solid black arrowheads shown in Fig. 26 indicate the sites of cleavage of the S-60 hairpin in the absence of the P-15 oligonucleotide and the hollow arrow heads indicate the sites of cleavage in the presence of the P-15 oligonucleotide. The size of the arrow head indicates the relative utilization of a particular site.

The S-60 hairpin molecule was labeled on its 5' end with biotin for subsequent detection. The S-60 hairpin was incubated in the presence of a thermostable 5' nuclease in the presence or the absence of the P-15 oligonucleotide. The presence of the full duplex which can be formed by the S-60 hairpin is demonstrated by cleavage with the Cleavase® BN 5' nuclease, in a primer-independent fashion (*i.e.*, in the absence of the P-15 oligonucleotide). The release of 18 and 19-nucleotide fragments from the 5' end of the S-60 hairpin molecule showed that the cleavage occurred near the junction between the single and double stranded regions when nothing is hybridized to the 3' arm of the S-60 hairpin (Fig. 27, lane 2).

The reactions shown in Fig. 27 were conducted as follows. Twenty fmole of the 5' biotin-labeled hairpin DNA (SEQ ID NO:29) was combined with 0.1 ng of Cleavase® BN

enzyme and 1 μl of 100 mM MOPS (pH 7.5) containing 0.5% each of Tween-20 and NP-40 in a total volume of 9 μl. In the reaction shown in lane 1, the enzyme was omitted and the volume was made up by addition of distilled water (this served as the uncut or no enzyme control). The reaction shown in lane 3 of Fig. 27 also included 0.5 pmole of the P15 oligonucleotide (SEQ ID NO:30), which can hybridize to the unpaired 3' arm of the S-60 hairpin (SEQ ID NO:29), as diagrammed in Fig. 26.

The reactions were overlaid with a drop of mineral oil, heated to 95°C for 15 seconds, then cooled to 37°C, and the reaction was started by the addition of 1 μl of 10 mM $MnCl_2$ to each tube. After 5 minutes, the reactions were stopped by the addition of 6 μl of 95% formamide containing 20 mM EDTA and 0.05% marker dyes. Samples were heated to 75°C for 2 minutes immediately before electrophoresis through a 15% acrylamide gel (19:1 cross-linked), with 7 M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA.

After electrophoresis, the gel plates were separated allowing the gel to remain flat on one plate. A 0.2 mm-pore positively-charged nylon membrane (NYTRAN, Schleicher and Schuell, Keene, NH), pre-wetted in H_2O, was laid on top of the exposed gel. All air bubbles were removed. Two pieces of 3MM filter paper (Whatman) were then placed on top of the membrane, the other glass plate was replaced, and the sandwich was clamped with binder clips. Transfer was allowed to proceed overnight. After transfer, the membrane was carefully peeled from the gel and allowed to air dry. After complete drying, the membrane was washed in 1.2X Sequenase Images Blocking Buffer (United States Biochemical) using 0.3 ml of buffer/cm^2 of membrane. The wash was performed for 30 minutes at room temperature. A streptavidin-alkaline phosphatase conjugate (SAAP, United States Biochemical) was added to a 1:4000 dilution directly to the blocking solution, and agitated for 15 minutes. The membrane was rinsed briefly with H_2O and then washed three times for 5 minutes per wash using 0.5 ml/cm^2 of 1X SAAP buffer (100 mM Tris-HCl, pH 10, 50 mM NaCl) with 0.1% sodium dodecyl sulfate (SDS). The membrane was rinsed briefly with H_20 between each wash. The membrane was then washed once in 1X SAAP buffer containing 1 mM $MgC1_2$ without SDS, drained thoroughly and placed in a plastic heat-sealable bag. Using a sterile pipet, 5 mls of CDP-Star™ (Tropix, Bedford, MA) chemiluminescent substrate for alkaline phosphatase were added to the bag and distributed over the entire membrane for 2-3 minutes. The CDP-Star™-treated membrane was exposed to XRP X-ray film (Kodak) for an initial exposure of 10 minutes.

The resulting autoradiograph is shown in Fig. 27. In Fig. 27, the lane labelled "M" contains the biotinylated P-15 oligonucleotide which served as a marker. The sizes (in nucleotides) of the uncleaved S-60 hairpin (60 nuc; lane 1), the marker (15 nuc; lane "M") and the cleavage products generated by cleavage of the S-60 hairpin in the presence (lane 3) or absence (lane 2) of the P-15 oligonucleotide are indicated.

Because the complementary regions of the S-60 hairpin are located on the same molecule, essentially no lag time should be needed to allow hybridization (*i.e.*, to form the duplex region of the hairpin). This hairpin structure would be expected to form long before the enzyme could locate and cleave the molecule. As expected, cleavage in the absence of the primer oligonucleotide was at or near the junction between the duplex and single-stranded regions, releasing the unpaired 5' arm (Fig. 27, lane 2). The resulting cleavage products were 18 and 19 nucleotides in length.

It was expected that stability of the S-60 hairpin with the tri-loop would prevent the P-15 oligonucleotide from promoting cleavage in the "primer-directed" manner described in Example 1 above, because the 3' end of the "primer" would remain unpaired. Surprisingly, it was found that the enzyme seemed to mediate an "invasion" by the P-15 primer into the duplex region of the S-60 hairpin, as evidenced by the shifting of the cleavage site 3 to 4 basepairs further into the duplex region, releasing the larger products (22 and 21 nuc.) observed in lane 3 of Fig. 27.

The precise sites of cleavage of the S-60 hairpin are diagrammed on the structure in Fig. 26, with the solid black arrowheads indicating the sites of cleavage in the absence of the P-15 oligonucleotide and the hollow arrow heads indicating the sites of cleavage in the presence of P-15.

These data show that the presence on the 3' arm of an oligonucleotide having some sequence homology with the first several bases of the similarly oriented strand of the downstream duplex can be a dominant factor in determining the site of cleavage by 5' nucleases. Because the oligonucleotide which shares some sequence homology with the first several bases of the similarly oriented strand of the downstream duplex appears to invade the duplex region of the hairpin, it is referred to as an" Invader™" oligonucleotide. As shown in the Examples below, an Invader™ oligonucleotide appears to invade (or displace) a region of duplexed nucleic acid regardless of whether the duplex region is present on the same molecule (*i.e.*, a hairpin) or whether the duplex is formed between two separate nucleic acid strands.

EXAMPLE 11

The Invader™ Oligonucleotide Shifts The Site Of Cleavage In A Pre-Formed Probe/Target Duplex

In Example 10, it was demonstrated that an Invader™ oligonucleotide could shift the site at which a 5' nuclease cleaves a duplex region present on a hairpin molecule. In this Example, the ability of an Invader™ oligonucleotide to shift the site of cleavage within a duplex region formed between two separate strands of nucleic acid molecules was examined.

A single-stranded target DNA comprising the single-stranded circular M13mp19 molecule and a labeled (fluorescein) probe oligonucleotide were mixed in the presence of the reaction buffer containing salt (KCl) and divalent cations (Mg^{2+} or Mn^{2+}) to promote duplex formation. The probe oligonucleotide refers to a labelled oligonucleotide which is complementary to a region along the target molecule (*e.g.*, M13mp19). A second oligonucleotide (unlabelled) was added to the reaction after the probe and target had been allowed to anneal. The second oligonucleotide binds to a region of the target which is located downstream of the region to which the probe oligonucleotide binds. This second oligonucleotide contains sequences which are complementary to a second region of the target molecule. If the second oligonucleotide contains a region which is complementary to a portion of the sequences along the target to which the probe oligonucleotide also binds, this second oligonucleotide is referred to as an Invader™ oligonucleotide (see Fig. 28c).

Fig. 32 depicts the annealing of two oligonucleotides to regions along the M13mp19 target molecule (bottom strand in all three structures shown). In Fig. 28 only a 52 nucleotide portion of the M13mp19 molecule is shown; this 52 nucleotide sequence is listed in SEQ ID NO:31. The probe oligonucleotide contains a fluorescein label at the 3' end; the sequence of the probe is 5'-AGAAAGGAAGGGAAGAAAGCGAAAGG-3' (SEQ ID NO:32). In Fig. 28, sequences comprising the second oligonucleotide, including the Invader™ oligonucleotide are underlined. In Fig. 28a, the second oligonucleotide, which has the sequence 5'-GACGGGGAAAGCCGGCGAACG-3' (SEQ ID NO:33), is complementary to a different and downstream region of the target molecule than is the probe oligonucleotide (labeled with fluorescein or "Fluor"); there is a gap between the second, upstream oligonucleotide and the probe for the structure shown in Fig. 28a. In Fig. 28b, the second, upstream oligonucleotide, which has the sequence 5'-GAAAGCCGGCGAACGTGGCG-3' (SEQ ID NO:34), is complementary to a different region of the target molecule than is the probe oligonucleotide, but in this case, the second oligonucleotide and the probe oligonucleotide abut one another

(that is the 3' end of the second, upstream oligonucleotide is immediately adjacent to the 5' end of the probe such that no gap exists between these two oligonucleotides). In Fig. 28c, the second, upstream oligonucleotide (5'-GGCGAACGTGGCGAGAAAGGA-3' [SEQ ID NO:35]) and the probe oligonucleotide share a region of complementarity with the target molecule. Thus, the upstream oligonucleotide has a 3' arm which has a sequence identical to the first several bases of the downstream probe. In this situation, the upstream oligonucleotide is referred to as an "Invader™" oligonucleotide.

The effect of the presence of an Invader™ oligonucleotide upon the pattern of cleavage in a probe/target duplex formed prior to the addition of the Invader™ was examined. The Invader™ oligonucleotide and the enzyme were added after the probe was allowed to anneal to the target and the position and extent of cleavage of the probe were examined to determine a) if the Invader™ was able to shift the cleavage site to a specific internal region of the probe, and b), if the reaction could accumulate specific cleavage products over time, even in the absence of thermal cycling, polymerization, or exonuclease removal of the probe sequence.

The reactions were carried out as follows. Twenty µl each of two enzyme mixtures were prepared, containing 2 µl of Cleavase® A/G nuclease extract (prepared as described in Example 2), with or without 50 pmole of the Invader™ oligonucleotide (SEQ ID NO:35), as indicated, per 4 µl of the mixture. For each of the eight reactions shown in Fig. 29, 150 fmole of M13mp19 single-stranded DNA (available from Life Technologies, Inc.) was combined with 5 pmoles of fluorescein labeled probe (SEQ ID NO:32), to create the structure shown in Fig. 28c, but without the Invader™ oligonucleotide present (the probe/target mixture). One half (4 tubes) of the probe/target mixtures were combined with 1 µl of 100 mM MOPS, pH 7.5 with 0.5% each of Tween-20 and NP-40, 0.5 µl of 1 M KCl and 0.25 µl of 80 mM $MnCl_2$, and distilled water to a volume of 6 µl. The second set of probe/target mixtures were combined with 1 µl of 100 mM MOPS, pH 7.5 with 0.5% each of Tween-20 and NP-40, 0.5 µl of 1 M KCl and 0.25 µl of 80 mM $MgCl_2$. The second set of mixtures therefore contained $MgCl_2$ in place of the $MnCl_2$ present in the first set of mixtures.

The mixtures (containing the probe/target with buffer, KCl and divalent cation) were covered with a drop of ChillOut® evaporation barrier and were brought to 60°C for 5 minutes to allow annealing. Four µl of the above enzyme mixtures without the Invader™ oligonucleotide was added to reactions whose products are shown in lanes 1, 3, 5 and 7 of Fig. 29. Reactions whose products are shown lanes 2, 4, 6, and 8 of Fig. 29 received the

same amount of enzyme mixed with the Invader™ oligonucleotide (SEQ ID NO:35). Reactions 1, 2, 5 and 6 were incubated for 5 minutes at 60°C and reactions 3, 4, 7 and 8 were incubated for 15 minutes at 60°C.

All reactions were stopped by the addition of 8 μl of 95% formamide with 20 mM EDTA and 0.05% marker dyes. Samples were heated to 90°C for 1 minute immediately before electrophoresis through a 20% acrylamide gel (19:1 cross-linked), containing 7 M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA. Following electrophoresis, the reaction products and were visualized by the use of an Hitachi FMBIO fluorescence imager, the output of which is seen in Fig. 29. The very low molecular weight fluorescent material seen in all lanes at or near the salt front in Fig. 29 and other fluoro-imager Figures is observed when fluorescently-labeled oligonucleotides are electrophoresed and imaged on a fluoro-imager. This material is not a product of the cleavage reaction.

The use of $MnCl_2$ in these reactions (lanes 1-4) stimulates the true exonuclease or "nibbling" activity of the Cleavase® enzyme, as described in Example 6, as is clearly seen in lanes 1 and 3 of Fig. 29. This nibbling of the probe oligonucleotide (SEQ ID NO:32) in the absence of Invader™ oligonucleotide (SEQ ID NO:35) confirms that the probe oligonucleotide is forming a duplex with the target sequence. The ladder-like products produced by this nibbling reaction may be difficult to differentiate from degradation of the probe by nucleases that might be present in a clinical specimen. In contrast, introduction of the Invader™ oligonucleotide (SEQ ID NO:35) caused a distinctive shift in the cleavage of the probe, pushing the site of cleavage 6 to 7 bases into the probe, confirming the annealing of both oligonucleotides. In presence of $MnCl_2$, the exonuclease "nibbling" may occur after the Invader™-directed cleavage event, until the residual duplex is destabilized and falls apart.

In a magnesium based cleavage reaction (lanes 5-8), the nibbling or true exonuclease function of the Cleavase® A/G is enzyme suppressed (but the endonucleolytic function of the enzyme is essentially unaltered), so the probe oligonucleotide is not degraded in the absence of the Invader™ (Fig. 29, lanes 5 and 7). When the Invader™ is added, it is clear that the Invader™ oligonucleotide can promote a shift in the site of the endonucleolytic cleavage of the annealed probe. Comparison of the products of the 5 and 15 minute reactions with Invader™ (lanes 6 and 8 in Fig. 29) shows that additional probe hybridizes to the target and is cleaved. The calculated melting temperature (T_m) of the portion of probe that is not invaded (*i.e.*, nucleotides 9-26 of SEQ ID NO:32) is 56°C, so the observed turnover (as evidenced by the accumulation of cleavage products with increasing reaction time) suggests

that the full length of the probe molecule, with a calculated T_m of 76°C, is must be involved in the subsequent probe annealing events in this 60°C reaction.

EXAMPLE 12

The Overlap Of The 3' Invader™ Oligonucleotide Sequence With The 5' Region Of The Probe Causes A Shift In The Site Of Cleavage

In Example 11, the ability of an Invader™ oligonucleotide to cause a shift in the site of cleavage of a probe annealed to a target molecule was demonstrated. In this Example, experiments were conducted to examine whether the presence of an oligonucleotide upstream from the probe was sufficient to cause a shift in the cleavage site(s) along the probe or whether the presence of nucleotides on the 3' end of the Invader™ oligonucleotide which have the same sequence as the first several nucleotides at the 5' end of the probe oligonucleotide were required to promote the shift in cleavage.

To examine this point, the products of cleavage obtained from three different arrangements of target-specific oligonucleotides are compared. A diagram of these oligonucleotides and the way in which they hybridize to a test nucleic acid, M13mp19, is shown in Fig. 28. In Fig. 28a, the 3' end of the upstream oligonucleotide (SEQ ID NO:33) is located upstream of the 5' end of the downstream "probe" oligonucleotide (SEQ ID NO:32) such that a region of the M13 target which is not paired to either oligonucleotide is present. In Fig. 28b, the sequence of the upstream oligonucleotide (SEQ ID NO:34) is immediately upstream of the probe (SEQ ID NO:32), having neither a gap nor an overlap between the sequences. Fig. 28c diagrams the arrangement of the substrates used in the assay of the present invention, showing that the upstream "Invader™" oligonucleotide (SEQ ID NO:35) has the same sequence on a portion of its 3' region as that present in the 5' region of the downstream probe (SEQ ID NO:32). That is to say, these regions will compete to hybridize to the same segment of the M13 target nucleic acid.

In these experiments, four enzyme mixtures were prepared as follows (planning 5 µl per digest): Mixture 1 contained 2.25µl of Cleavase® A/G nuclease extract (prepared as described in Example 2) per 5 µl of mixture, in 20 mM MOPS, pH 7.5 with 0.1 % each of Tween 20 and NP-40, 4 mM $MnCl_2$ and 100 mM KCl. Mixture 2 contained 11.25 units of *Taq* DNA polymerase (Promega) per 5 µl of mixture in 20 mM MOPS, pH 7.5 with 0.1 % each of Tween 20 and NP-40, 4 mM $MnCl_2$ and 100 mM KCl. Mixture 3 contained 2.25 µl of Cleavase® A/G nuclease extract per 5 µl of mixture in 20 mM Tris-HCl, pH 8.5, 4 mM

$MgCl_2$ and 100 mM KCl. Mixture 4 contained 11.25 units of *Taq* DNA polymerase per 5 µl of mixture in 20 mM Tris-HCl, pH 8.5, 4 mM $MgCl_2$ and 100 mM KCl.

For each reaction, 50 fmole of M13mp19 single-stranded DNA (the target nucleic acid) was combined with 5 pmole of the probe oligonucleotide (SEQ ID NO:32 which contained a fluorescein label at the 3' end) and 50 pmole of one of the three upstream oligonucleotides diagrammed in Fig. 28 (*i.e.*, one of SEQ ID NOS:33-35), in a total volume of 5 µl of distilled water. The reactions were overlaid with a drop of ChillOut™ evaporation barrier and warmed to 62°C. The cleavage reactions were started by the addition of 5 µl of an enzyme mixture to each tube, and the reactions were incubated at 62°C for 30 min. The reactions shown in lanes 1-3 of Fig. 30 received Mixture 1; reactions 4-6 received Mixture 2; reactions 7-9 received Mixture 3 and reactions 10-12 received Mixture 4.

After 30 minutes at 62°C, the reactions were stopped by the addition of 8 µl of 95% formamide with 20 mM EDTA and 0.05% marker dyes. Samples were heated to 75°C for 2 minutes immediately before electrophoresis through a 20% acrylamide gel (19:1 cross-linked), with 7 M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA.

Following electrophoresis, the products of the reactions were visualized by the use of an Hitachi FMBIO fluorescence imager, the output of which is seen in Fig. 30. The reaction products shown in lanes 1, 4, 7 and 10 of Fig. 30 were from reactions which contained SEQ ID NO:33 as the upstream oligonucleotide (see Fig. 28a). The reaction products shown in lanes 2, 5, 8 and 11 of Fig. 30 were from reactions which contained SEQ ID NO:34 as the upstream oligonucleotide (see Fig. 28b). The reaction products shown in lanes 3, 6, 9 and 12 of Fig. 30 were from reactions which contained SEQ ID NO:35, the Invader™ oligonucleotide, as the upstream oligonucleotide (see Fig. 28c).

Examination of the Mn^{2+} based reactions using either Cleavase® A/G nuclease or DNAP*Taq* as the cleavage agent (lanes 1 through 3 and 4 through 6, respectively) shows that both enzymes have active exonuclease function in these buffer conditions. The use of a 3' label on the probe oligonucleotide allows the products of the nibbling activity to remain labeled, and therefore visible in this assay. The ladders seen in lanes 1, 2, 4 and 5 confirm that the probe hybridize to the target DNA as intended. These lanes also show that the location of the non-invasive oligonucleotides have little effect on the products generated. The uniform ladder created by these digests would be difficult to distinguish from a ladder causes by a contaminating nuclease, as one might find in a clinical specimen. In contrast, the products displayed in lanes 3 and 6, where an Invader™ oligonucleotide was provided to

direct the cleavage, show a very distinctive shift, so that the primary cleavage product is smaller than those seen in the non-invasive cleavage. This product is then subject to further nibbling in these conditions, as indicated by the shorter products in these lanes. These Invader™-directed cleavage products would be easily distinguished from a background of non-specific degradation of the probe oligonucleotide.

When Mg^{2+} is used as the divalent cation the results are even more distinctive. In lanes 7, 8, 10 and 11 of Fig. 30, where the upstream oligonucleotides were not invasive, minimal nibbling is observed. The products in the DNAP*Taq* reactions show some accumulation of probe that has been shortened on the 5' end by one or two nucleotides consistent with previous examination of the action of this enzyme on nicked substrates (Longley *et al.*, *supra*). When the upstream oligonucleotide is invasive, however, the appearance of the distinctively shifted probe band is seen. These data clearly indicated that it is the invasive 3' portion of the upstream oligonucleotide that is responsible for fixing the site of cleavage of the downstream probe.

Thus, the above results demonstrate that it is the presence of the free or initially non-annealed nucleotides at the 3' end of the Invader™ oligonucleotide which mediate the shift in the cleavage site, not just the presence of an oligonucleotide annealed upstream of the probe. Nucleic acid detection assays which employ the use of an Invader™ oligonucleotide are termed "Invader™-directed cleavage" assays.

EXAMPLE 13

Invader™-Directed Cleavage Recognizes Single And Double Stranded Target Molecules In A Background Of Non-Target DNA Molecules

For a nucleic acid detection method to be broadly useful, it must be able to detect a specific target in a sample that may contain large amounts of other DNA, (*e.g.*, bacterial or human chromosomal DNA). The ability of the Invader™ directed cleavage assay to recognize and cleave either single- or double-stranded target molecules in the presence of large amounts of non-target DNA was examined. In these experiments a model target nucleic acid, M13, in either single or double stranded form (single-stranded M13mp18 is available from Life Technologies, Inc and double-stranded M13mp19 is available from NEB), was combined with human genomic DNA (Novagen) and then utilized in Invader™-directed cleavage reactions. Before the start of the cleavage reaction, the DNAs were heated to 95°C for 15 minutes to completely denature the samples, as is standard practice in assays, such as

polymerase chain reaction or enzymatic DNA sequencing, which involve solution hybridization of oligonucleotides to double-stranded target molecules.

For each of the reactions shown in lanes 2-5 of Fig. 31, the target DNA (25 fmole of the ss DNA or 1 pmole of the ds DNA) was combined with 50 pmole of the Invader™ oligonucleotide (SEQ ID NO:35); for the reaction shown in lane 1 the target DNA was omitted. Reactions 1, 3 and 5 also contained 470 ng of human genomic DNA. These mixtures were brought to a volume of 10 μl with distilled water, overlaid with a drop of ChillOut™ evaporation barrier, and brought to 95°C for 15 minutes. After this incubation period, and still at 95°C, each tube received 10 μl of a mixture comprising 2.25 μl of Cleavase® A/G nuclease extract (prepared as described in Example 2) and 5 pmole of the probe oligonucleotide (SEQ ID NO:32), in 20 mM MOPS, pH 7.5 with 0.1 % each of Tween 20 and NP-40, 4 mM $MnCl_2$ and 100 mM KCl. The reactions were brought to 62°C for 15 minutes and stopped by the addition of 12 μl of 95% formamide with 20 mM EDTA and 0.05% marker dyes. Samples were heated to 75°C for 2 minutes immediately before electrophoresis through a 20% acrylamide gel (19:1 cross-linked), with 7 M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA. The products of the reactions were visualized by the use of an Hitachi FMBIO fluorescence imager. The results are displayed in Fig. 31.

In Fig. 31, lane 1 contains the products of the reaction containing the probe (SEQ ID NO:32), the Invader™ oligonucleotide (SEQ ID NO:35) and human genomic DNA. Examination of lane 1 shows that the probe and Invader™ oligonucleotides are specific for the target sequence, and that the presence of genomic DNA does not cause any significant background cleavage.

In Fig. 31, lanes 2 and 3 contain reaction products from reactions containing the single-stranded target DNA (M13mp18), the probe (SEQ ID NO:32) and the Invader™ oligonucleotide (SEQ ID NO:35) in the absence or presence of human genomic DNA, respectively. Examination of lanes 2 and 3 demonstrate that the Invader™ detection assay may be used to detect the presence of a specific sequence on a single-stranded target molecule in the presence or absence of a large excess of competitor DNA (human genomic DNA).

In Fig. 31, lanes 4 and 5 contain reaction products from reactions containing the double-stranded target DNA (M13mp19), the probe (SEQ ID NO:32) and the Invader™ oligonucleotide (SEQ ID NO:35) in the absence or presence of human genomic DNA, respectively. Examination of lanes 4 and 5 show that double stranded target molecules are

eminently suitable for Invader™-directed detection reactions. The success of this reaction using a short duplexed molecule, M13mp19, as the target in a background of a large excess of genomic DNA is especially noteworthy as it would be anticipated that the shorter and less complex M13 DNA strands would be expected to find their complementary strand more easily than would the strands of the more complex human genomic DNA. If the M13 DNA reannealed before the probe and/or Invader™ oligonucleotides could bind to the target sequences along the M13 DNA, the cleavage reaction would be prevented. In addition, because the denatured genomic DNA would potentially contain regions complementary to the probe and/or Invader™ oligonucleotides it was possible that the presence of the genomic DNA would inhibit the reaction by binding these oligonucleotides thereby preventing their hybridization to the M13 target. The above results demonstrate that these theoretical concerns are not a problem under the reaction conditions employed above.

In addition to demonstrating that the Invader™ detection assay may be used to detect sequences present in a double-stranded target, these data also show that the presence of a large amount of non-target DNA (470 ng/20 μl reaction) does not lessen the specificity of the cleavage. While this amount of DNA does show some impact on the rate of product accumulation, probably by binding a portion of the enzyme, the nature of the target sequence, whether single- or double-stranded nucleic acid, does not limit the application of this assay.

EXAMPLE 14

Signal Accumulation In The Invader™-Directed Cleavage Assay As A Function Of Target Concentration

To investigate whether the Invader™-directed cleavage assay could be used to indicate the amount of target nucleic acid in a sample, the following experiment was performed. Cleavage reactions were assembled which contained an Invader™ oligonucleotide (SEQ ID NO:35), a labelled probe (SEQ ID NO:32) and a target nucleic acid, M13mp19. A series of reactions, which contained smaller and smaller amounts of the M13 target DNA, was employed in order to examine whether the cleavage products would accumulate in a manner that reflected the amount of target DNA present in the reaction.

The reactions were conducted as follows. A master mix containing enzyme and buffer was assembled. Each 5 μl of the master mixture contained 25 ng of Cleavase® BN nuclease in 20 mM MOPS (pH 7.5) with 0.1% each of Tween 20 and NP-40, 4 mM $MnCl_2$ and 100 mM KCl. For each of the cleavage reactions shown in lanes 4-13 of Fig. 32, a DNA mixture

was generated which contained 5 pmoles of the fluorescein-labelled probe oligonucleotide (SEQ ID NO:32), 50 pmoles of the Invader™ oligonucleotide (SEQ ID NO:35) and 100, 50, 10, 5, 1, 0.5, 0.1, 0.05, 0.01 or 0.005 fmoles of single-stranded M13mp19, respectively, for every 5 μl of the DNA mixture. The DNA solutions were covered with a drop of ChillOut® evaporation barrier and brought to 61°C. The cleavage reactions were started by the addition of 5 μl of the enzyme mixture to each of tubes (final reaction volume was 10 μl). After 30 minutes at 61°C, the reactions were terminated by the addition of 8 μl of 95% formamide with 20 mM EDTA and 0.05% marker dyes. Samples were heated to 90°C for 1 minute immediately before electrophoresis through a 20% denaturing acrylamide gel (19:1 cross-linked) with 7 M urea, in a buffer containing 45 mM Tris-Borate (pH 8.3), 1.4 mM EDTA. To provide reference (*i.e.*, standards), 1.0, 0.1 and 0.01 pmole aliquots of fluorescein-labelled probe oligonucleotide (SEQ ID NO:32) were diluted with the above formamide solution to a final volume of 18 μl. These reference markers were loaded into lanes 1-3, respectively of the gel. The products of the cleavage reactions (as well as the reference standards) were visualized following electrophoresis by the use of a Hitachi FMBIO fluorescence imager. The results are displayed in Fig. 32.

In Fig. 32, boxes appear around fluorescein-containing nucleic acid (*i.e.*, the cleaved and uncleaved probe molecules) and the amount of fluorescein contained within each box is indicated under the box. The background fluorescence of the gel (see box labelled "background") was subtracted by the fluoro-imager to generate each value displayed under a box containing cleaved or uncleaved probe products (the boxes are numbered 1-14 at top left with a V followed by a number below the box). The lane marked "M" contains fluoresceinated oligonucleotides which served as markers.

The results shown in Fig. 32, demonstrate that the accumulation of cleaved probe molecules in a fixed-length incubation period reflects the amount of target DNA present in the reaction. The results also demonstrate that the cleaved probe products accumulate in excess of the copy number of the target. This is clearly demonstrated by comparing the results shown in lane 3, in which 10 fmole (0.01 pmole) of uncut probe are displayed with the results shown in 5, where the products which accumulated in response to the presence of 10 fmole of target DNA are displayed. These results show that the reaction can cleave hundreds of probe oligonucleotide molecules for each target molecule present, dramatically amplifying the target-specific signal generated in the Invader™-directed cleavage reaction.

EXAMPLE 15

Effect Of Saliva Extract On The Invader™-Directed Cleavage Assay

For a nucleic acid detection method to be useful in a medical (*i.e.*, a diagnostic) setting, it must not be inhibited by materials and contaminants likely to be found in a typical clinical specimen. To test the susceptibility of the Invader™-directed cleavage assay to various materials, including but not limited to nucleic acids, glycoproteins and carbohydrates, likely to be found in a clinical sample, a sample of human saliva was prepared in a manner consistent with practices in the clinical laboratory and the resulting saliva extract was added to the Invader™-directed cleavage assay. The effect of the saliva extract upon the inhibition of cleavage and upon the specificity of the cleavage reaction was examined.

One and one-half milliliters of human saliva were collected and extracted once with an equal volume of a mixture containing phenol:chloroform:isoamyl alcohol (25:24:1). The resulting mixture was centrifuged in a microcentrifuge to separate the aqueous and organic phases. The upper, aqueous phase was transferred to a fresh tube. One-tenth volumes of 3 M NaOAc were added and the contents of the tube were mixed. Two volumes of 100% ethyl alcohol were added to the mixture and the sample was mixed and incubated at room temperature for 15 minutes to allow a precipitate to form. The sample was centrifuged in a microcentrifuge at 13,000 rpm for 5 minutes and the supernatant was removed and discarded. A milky pellet was easily visible. The pellet was rinsed once with 70% ethanol, dried under vacuum and dissolved in 200 μl of 10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA (this constitutes the saliva extract). Each μl of the saliva extract was equivalent to 7.5 μl of saliva. Analysis of the saliva extract by scanning ultraviolet spectrophotometry showed a peak absorbance at about 260 nm and indicated the presence of approximately 45 ng of total nucleic acid per μl of extract.

The effect of the presence of saliva extract upon the following enzymes was examined: Cleavase® BN nuclease, Cleavase® A/G nuclease and three different lots of DNAPTaq: AmpliTaq® (Perkin Elmer; a recombinant form of DNAPTaq), AmpliTaq® LD (Perkin-Elmer; a recombinant DNAPTaq preparation containing very low levels of DNA) and *Taq* DNA polymerase (Fischer). For each enzyme tested, an enzyme/probe mixture was made comprising the chosen amount of enzyme with 5 pmole of the probe oligonucleotide (SEQ ID NO:32) in 10 μl of 20 mM MOPS (pH 7.5) containing 0.1% each of Tween 20 and NP-40, 4 mM $MnCl_2$, 100 mM KCl and 100 μg/ml BSA. The following amounts of enzyme were used: 25 ng of Cleavase® BN prepared as described in Example 8; 2 μl of Cleavase® A/G

nuclease extract prepared as described in Example 2; 2.25 μl (11.25 polymerase units) the following DNA polymerases: AmpliTaq® DNA polymerase (Perkin Elmer); AmpliTaq® DNA polymerase LD (low DNA; from Perkin Elmer); *Taq* DNA polymerase (Fisher Scientific).

For each of the reactions shown in Fig. 33, except for that shown in lane 1, the target DNA (50 fmoles of single-stranded M13mp19 DNA) was combined with 50 pmole of the Invader™ oligonucleotide (SEQ ID NO:35) and 5 pmole of the probe oligonucleotide (SEQ ID NO:32); target DNA was omitted in reaction 1 (lane 1). Reactions 1, 3, 5, 7, 9 and 11 included 1.5 μl of saliva extract. These mixtures were brought to a volume of 5 μl with distilled water, overlaid with a drop of ChillOut® evaporation barrier and brought to 95°C for 10 minutes. The cleavage reactions were then started by the addition of 5 μl of the desired enzyme/probe mixture; reactions 1, 4 and 5 received Cleavase® A/G nuclease. Reactions 2 and 3 received Cleavase® BN; reactions 6 and 7 received AmpliTaq®; reactions 8 and 9 received AmpliTaq® LD; and reactions 10 and 11 received *Taq* DNA Polymerase from Fisher Scientific.

The reactions were incubated at 63°C for 30 minutes and were stopped by the addition of 6 μl of 95% formamide with 20 mM EDTA and 0.05% marker dyes. Samples were heated to 75°C for 2 minutes immediately before electrophoresis through a 20% acrylamide gel (19:1 cross-linked), with 7 M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA. The products of the reactions were visualized by the use of an Hitachi FMBIO fluorescence imager, and the results are displayed in Fig. 33.

A pairwise comparison of the lanes shown in Fig. 33 without and with the saliva extract, treated with each of the enzymes, shows that the saliva extract has different effects on each of the enzymes. While the Cleavase® BN nuclease and the AmpliTaq® are significantly inhibited from cleaving in these conditions, the Cleavase® A/G nuclease and AmpliTaq® LD display little difference in the yield of cleaved probe. The preparation of *Taq* DNA polymerase from Fisher Scientific shows an intermediate response, with a partial reduction in the yield of cleaved product. From the standpoint of polymerization, the three DNAPTaq variants should be equivalent; these should be the same protein with the same amount of synthetic activity. It is possible that the differences observed could be due to variations in the amount of nuclease activity present in each preparation caused by different handling during purification, or by different purification protocols. In any case, quality control assays designed to assess polymerization activity in commercial DNAP preparations would be

unlikely to reveal variation in the amount of nuclease activity present. If preparations of DNAPTaq were screened for full 5' nuclease activity (*i.e.*, if the 5' nuclease activity was specifically quantitated), it is likely that the preparations would display sensitivities (to saliva extract) more in line with that observed using Cleavase® A/G nuclease, from which DNAPTaq differs by a very few amino acids.

It is worthy of note that even in the slowed reactions of Cleavase® BN and the DNAPTaq variants there is no noticeable increase in non-specific cleavage of the probe oligonucleotide due to inappropriate hybridization or saliva-borne nucleases.

EXAMPLE 16

Comparison Of Additional 5' Nucleases In The Invader™-Directed Cleavage Assay

A number of eubacterial Type A DNA polymerases (*i.e.*, Pol I type DNA polymerases) have been shown to function as structure specific endonucleases (*See*, Example 1, and Lyamichev *et al.*, *supra*). In this Example, it was demonstrated that the enzymes of this class can also be made to catalyze the Invader™-directed cleavage of the present invention, albeit not as efficiently as the Cleavase® enzymes.

Cleavase® BN nuclease and Cleavase® A/G nuclease were tested along side three different thermostable DNA polymerases: *Thermus aquaticus* DNA polymerase (Promega), *Thermus thermophilus* and *Thermus flavus* DNA polymerases (Epicentre). The enzyme mixtures used in the reactions shown in lanes 1-11 of Fig. 34 contained the following, each in a volume of 5 µl: Lane 1: 20 mM MOPS (pH 7.5) with 0.1% each of Tween 20 and NP-40, 4 mM $MnCl_2$, 100 mM KCl; Lane 2: 25 ng of Cleavase® BN nuclease in the same solution described for lane 1; Lane 3: 2.25 µl of Cleavase® A/G nuclease extract (prepared as described in Example 2), in the same solution described for lane 1; Lane 4: 2.25 µl of Cleavase® A/G nuclease extract in 20 mM Tris-Cl, (pH 8.5), 4 mM $MgCl_2$ and 100 mM KCl; Lane 5: 11.25 polymerase units of *Taq* DNA polymerase in the same buffer described for lane 4; Lane 6: 11.25 polymerase units of *Tth* DNA polymerase in the same buffer described for lane 1; Lane 7: 11.25 polymerase units of *Tth* DNA polymerase in a 2X concentration of the buffer supplied by the manufacturer, supplemented with 4 mM $MnCl_2$; Lane 8: 11.25 polymerase units of *Tth* DNA polymerase in a 2X concentration of the buffer supplied by the manufacturer, supplemented with 4 mM $MgCl_2$; Lane 9: 2.25 polymerase units of *Tfl* DNA polymerase in the same buffer described for lane 1; Lane 10: 2.25

polymerase units of Tfl polymerase in a 2X concentration of the buffer supplied by the manufacturer, supplemented with 4 mM $MnCl_2$; Lane 11: 2.25 polymerase units of *Tfl* DNA polymerase in a 2X concentration of the buffer supplied by the manufacturer, supplemented with 4 mM $MgCl_2$.

Sufficient target DNA, probe and Invader™ for all 11 reactions was combined into a master mix. This mix contained 550 fmoles of single-stranded Ml3mp19 target DNA, 550 pmoles of the Invader™ oligonucleotide (SEQ ID NO:35) and 55 pmoles of the probe oligonucleotide (SEQ ID NO:32), each as depicted in Fig. 28c, in 55 µl of distilled water. Five µl of the DNA mixture was dispensed into each of 11 labeled tubes and overlaid with a drop of ChillOut® evaporation barrier. The reactions were brought to 63°C and cleavage was started by the addition of 5 µl of the appropriate enzyme mixture. The reaction mixtures were then incubated at 63°C temperature for 15 minutes. The reactions were stopped by the addition of 8 µl of 95% formamide with 20 mM EDTA and 0.05% marker dyes. Samples were heated to 90°C for 1 minute immediately before electrophoresis through a 20% acrylamide gel (19:1 cross-linked), with 7 M urea, in a buffer of 45 mM Tris-Borate (pH 8.3), 1.4 mM EDTA. Following electrophoresis, the products of the reactions were visualized by the use of an Hitachi FMBIO fluorescence imager, and the results are displayed in Fig. 34. Examination of the results shown in Fig. 34 demonstrates that all of the 5' nucleases tested have the ability to catalyze Invader™-directed cleavage in at least one of the buffer systems tested. Although not optimized here, these cleavage agents are suitable for use in the methods of the present invention.

EXAMPLE 17

The Invader™-Directed Cleavage Assay Can Detect Single Base Differences In Target Nucleic Acid Sequences

The ability of the Invader™-directed cleavage assay to detect single base mismatch mutations was examined. Two target nucleic acid sequences containing Cleavase® enzyme-resistant phosphorothioate backbones were chemically synthesized and purified by polyacrylamide gel electrophoresis. Targets comprising phosphorothioate backbones were used to prevent exonucleolytic nibbling of the target when duplexed with an oligonucleotide. A target oligonucleotide, which provides a target sequence that is completely complementary to the Invader™ oligonucleotide (SEQ ID NO:35) and the probe oligonucleotide (SEQ ID NO:32), contained the following sequence:

5'-CCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGC-3' (SEQ ID NO:36). A second target sequence containing a single base change relative to SEQ ID NO:36 was synthesized: 5'-CCTTTCGCT**C**TCTTCCCTTCCTTTCTCGCC ACGTTCGCCGGC-3 (SEQ ID NO:37; the single base change relative to SEQ ID NO:36 is shown using bold and underlined type). The consequent mismatch occurs within the "Z" region of the target as represented in Fig. 25.

To discriminate between two target sequences which differ by the presence of a single mismatch), Invader™-directed cleavage reactions were conducted using two different reaction temperatures (55°C and 60°C). Mixtures containing 200 fmoles of either SEQ ID NO:36 or SEQ ID NO:37, 3 pmoles of fluorescein-labelled probe oligonucleotide (SEQ ID NO:32), 7.7 pmoles of Invader™ oligonucleotide (SEQ ID NO:35) and 2 μl of Cleavase® A/G nuclease extract (prepared as described in Example 2) in 9 μl of 10 mM MOPS (pH 7.4) with 50 mM KCl were assembled, covered with a drop of ChillOut® evaporation barrier and brought to the appropriate reaction temperature. The cleavage reactions were initiated by the addition of 1 μl of 20 mM $MgCl_2$. After 30 minutes at either 55°C or 60°C, 10 μl of 95% formamide with 20 mM EDTA and 0.05% marker dyes was added to stop the reactions. The reaction mixtures where then heated to 90°C for one minute prior to loading 4 μl onto 20% denaturing polyacrylamide gels. The resolved reaction products were visualized using a Hitachi FMBIO fluorescence imager. The resulting image is shown in Fig. 35.

In Fig. 35, lanes 1 and 2 show the products from reactions conducted at 55°C; lanes 3 and 4 show the products from reactions conducted at 60°C. Lanes 1 and 3 contained products from reactions containing SEQ ID NO:36 (perfect match to probe) as the target. Lanes 2 and 4 contained products from reactions containing SEQ ID NO:37 (single base mismatch with probe) as the target. The target that does not have a perfect hybridization match (*i.e.*, complete complementarity) with the probe will not bind as strongly (*i.e.*, the T_m of that duplex will be lower than the T_m of the same region if perfectly matched). The results presented here show that reaction conditions can be varied to either accommodate the mismatch (*e.g.*, by lowering the temperature of the reaction) or to exclude the binding of the mismatched sequence (*e.g.*, by raising the reaction temperature).

The results shown in Fig. 35 demonstrate that the specific cleavage event which occurs in Invader™-directed cleavage reactions can be eliminated by the presence of a single base mis-match between the probe oligonucleotide and the target sequence. Thus, reaction conditions can be chosen so as to exclude the hybridization of mis-matched Invader™-

directed cleavage probes thereby diminishing or even eliminating the cleavage of the probe. In an extension of this assay system, multiple cleavage probes, each possessing a separate reporter molecule (*i.e.*, a unique label), could also be used in a single cleavage reaction, to simultaneously probe for two or more variants in the same target region. The products of such a reaction would allow not only the detection of mutations which exist within a target molecule, but would also allow a determination of the relative concentrations of each sequence (*i.e.*, mutant and wild type or multiple different mutants) present within samples containing a mixture of target sequences. When provided in equal amounts, but in a vast excess (*e.g.*, at least a 100-fold molar excess; typically at least 1 pmole of each probe oligonucleotide would be used when the target sequence was present at about 10 fmoles or less) over the target and used in optimized conditions. As discussed above, any differences in the relative amounts of the target variants will not affect the kinetics of hybridization, so the amounts of cleavage of each probe will reflect the relative amounts of each variant present in the reaction.

The results shown in the Example clearly demonstrate that the Invader™-directed cleavage reaction can be used to detect single base difference between target nucleic acids.

EXAMPLE 18

The Invader™-Directed Cleavage Reaction Is Insensitive To Large Changes In Reaction Conditions

The results shown above demonstrated that the Invader™-directed cleavage reaction can be used for the detection of target nucleic acid sequences and that this assay can be used to detect single base difference between target nucleic acids. These results demonstrated that 5' nucleases (*e.g.*, Cleavase®BN, Cleavase® A/G, DNAPTaq, DNAPTth, DNAPTfl) could be used in conjunction with a pair of overlapping oligonucleotides as an efficient way to recognize nucleic acid targets. In the experiments below it is demonstrated that invasive cleavage reaction is relatively insensitive to large changes in conditions thereby making the method suitable for practice in clinical laboratories.

The effects of varying the conditions of the cleavage reaction were examined for their effect(s) on the specificity of the invasive cleavage and the on the amount of signal accumulated in the course of the reaction. To compare variations in the cleavage reaction a "standard" Invader™ cleavage reaction was first defined. In each instance, unless specifically stated to be otherwise, the indicated parameter of the reaction was varied, while the invariant

aspects of a particular test were those of this standard reaction. The results of these tests are either shown in Figs. 38-40, or the results described below.

a) The Standard Invader™-Directed Cleavage Reaction

The standard reaction was defined as comprising 1 fmole of M13mp18 single-stranded target DNA (NEB), 5 pmoles of the labeled probe oligonucleotide (SEQ ID NO:38), 10 pmole of the upstream Invader™ oligonucleotide (SEQ ID NO:39) and 2 units of Cleavase® A/G in 10 μl of 10 mM MOPS, pH 7.5 with 100 mM KCl, 4 mM $MnCl_2$, and 0.05% each Tween-20 and Nonidet-P40. For each reaction, the buffers, salts and enzyme were combined in a volume of 5 μl; the DNAs (target and two oligonucleotides) were combined in 5 μl of dH_2O and overlaid with a drop of ChillOut® evaporation barrier. When multiple reactions were performed with the same reaction constituents, these formulations were expanded proportionally.

Unless otherwise stated, the sample tubes with the DNA mixtures were warmed to 61°C, and the reactions were started by the addition of 5 μl of the enzyme mixture. After 20 minutes at this temperature, the reactions were stopped by the addition of 8 μl of 95% formamide with 20 mM EDTA and 0.05% marker dyes. Samples were heated to 75°C for 2 minutes immediately before electrophoresis through a 20% acrylamide gel (19:1 cross-linked), with 7 M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA. The products of the reactions were visualized by the use of an Hitachi FMBIO fluorescence imager. In each case, the uncut probe material was visible as an intense black band or blob, usually in the top half of the panel, while the desired products of Invader™ specific cleavage were visible as one or two narrower black bands, usually in the bottom half of the panel. Under some reaction conditions, particularly those with elevated salt concentrations, a secondary cleavage product is also visible (thus generating a doublet). Ladders of lighter grey bands generally indicate either exonuclease nibbling of the probe oligonucleotide or heat-induced, non-specific breakage of the probe.

Fig. 37 depicts the annealing of the probe and Invader™ oligonucleotides to regions along the M13mp18 target molecule (the bottom strand). In Fig. 37 only a 52 nucleotide portion of the M13mp18 molecule is shown; this 52 nucleotide sequence is listed in SEQ ID NO:31 (this sequence is identical in both M13mp18 and M13mp19). The probe oligonucleotide (top strand) contains a Cy3 amidite label at the 5' end; the sequence of the probe is 5'-AGAAAGGAAGGGAAGAAAGCGAAAGGT-3' (SEQ ID NO:38. The bold

type indicates the presence of a modified base (2'-O-CH_3). Cy3 amidite (Pharmacia) is a indodicarbocyanine dye amidite which can be incorporated at any position during the synthesis of oligonucleotides; Cy3 fluoresces in the yellow region (excitation and emission maximum of 554 and 568 nm, respectively). The Invader™ oligonucleotide (middle strand) has the following sequence: 5'-GCCGGCGAACGTGGCGAGAAAGGA-3' (SEQ ID NO:39).

b) KCl Titration

Fig. 38 shows the results of varying the KCl concentration in combination with the use of 2 mM $MnCl_2$, in an otherwise standard reaction. The reactions were performed in duplicate for confirmation of observations; the reactions shown in lanes 1 and 2 contained no added KCl, lanes 3 and 4 contained KCl at 5 mM, lanes 5 and 6 contained 25 mM KCl, lanes 7 and 8 contained 50 mM KCl, lanes 9 and 10 contained 100 mM KCl and lanes 11 and 12 contained 200 mM KCl. These results show that the inclusion of KCl allows the generation of a specific cleavage product. While the strongest signal is observed at the 100 mM KCl concentration, the specificity of signal in the other reactions with KCl at or above 25 mM indicates that concentrations in the full range (*i.e.*, 25-200 mM) may be chosen if it is so desirable for any particular reaction conditions.

As shown in Fig. 38, the Invader™-directed cleavage reaction requires the presence of salt (*e.g.*, KCl) for effective cleavage to occur. In other reactions, it has been found that KCl can inhibit the activity of certain Cleavase® enzymes when present at concentrations above about 25 mM. For example, in cleavage reactions using the S-60 oligonucleotide shown in Fig. 26, in the absence of primer, the Cleavase® BN enzyme loses approximately 50% of its activity in 50 mM KCl. Therefore, the use of alternative salts in the Invader™-directed cleavage reaction was examined. In these experiments, the potassium ion was replaced with either Na^+ or Li^+ or the chloride ion was replaced with glutamic acid. The replacement of KCl with alternative salts is described below in Sections c-e.

c) NaCl Titration

NaCl was used in place of KCl at 75, 100, 150 or 200 mM, in combination with the use 2 mM $MnCl_2$, in an otherwise standard reaction. These results showed that NaCl can be used as a replacement for KCl in the Invader™-directed cleavage reaction, with like

concentration giving like results, (*i.e.*, the presence of NaCl, like KCl, enhances product accumulation).

d) LiCl Titration

LiCl was used in place of KCl in otherwise standard reactions. Concentrations tested were 25, 50, 75, 100, 150 and 200mM LiCl. The results demonstrated that LiCl can be used as a suitable replacement for KCl in the Invader™-directed cleavage reaction (*i.e.*, the presence of LiCl, like KCl, enhances product accumulation), in concentrations of about 100 mM or higher.

e) KGlu Titration

The results of using a glutamate salt of potassium (KGlu) in place of the more commonly used chloride salt (KCl) in reactions performed over a range of temperatures were examined. KGlu has been shown to be a highly effective salt source for some enzymatic reactions, showing a broader range of concentrations which permit maximum enzymatic activity (Leirmo *et al.*, Biochem., 26:2095 [1987]). The ability of KGlu to facilitate the annealing of the probe and Invader™ oligonucleotides to the target nucleic acid was compared to that of LiCl. In these experiments, the reactions were run for 15 minutes, rather than the standard 20 minutes, in standard reactions that replaced KCl 200 mM, 300 mM or 400 mM KGlu. The reactions were run at 65°C, 67°C, 69°C or 71°C. The results showed demonstrated that KGlu was very effective as a salt in the invasive cleavage reactions, with full activity apparent even at 400 mM KGlu, though at the lowest temperature cleavage was reduced by about 30% at 300 mM KGlu, and by about 90% to 400 mM KGlu.

f) $MnCl_2$ And $MgCl_2$ Titration And Ability To Replace $MnCl_2$ With $MgCl_2$

In some instances it may be desirable to perform the invasive cleavage reaction in the presence of Mg^{2+}, either in addition to, or in place of Mn^{2+} as the necessary divalent cation required for activity of the enzyme employed. For example, some common methods of preparing DNA from bacterial cultures or tissues use $MgCl_2$ in solutions which are used to facilitate the collection of DNA by precipitation. In addition, elevated concentrations (*i.e.*, greater than 5 mM) of divalent cation can be used to facilitate hybridization of nucleic acids, in the same way that the monovalent salts were used above, thereby enhancing the invasive

cleavage reaction. In this experiment, the tolerance of the invasive cleavage reaction was examined for 1) the substitution of $MgCl_2$ for $MnCl_2$ and for the ability to produce specific product in the presence of increasing concentrations of $MgCl_2$ and $MnCl_2$.

Fig. 39 shows the results of either varying the concentration of $MnCl_2$ from 2 mM to 8 mM, replacing the $MnCl_2$ with $MgCl_2$ at 2 to 4 mM, or of using these components in combination in an otherwise standard reaction. The reactions analyzed in lanes 1 and 2 contained 2 mM each $MnCl_2$ and $MgCl_2$, lanes 3 and 4 contained 2 mM $MnCl_2$ only, lanes 5 and 6 contained 3 mM $MnCl_2$, lanes 7 and 8 contained 4 mM $MnCl_2$, lanes 9 and 10 contained 8 mM $MnCl_2$. The reactions analyzed in lanes 11 and 12 contained 2 mM $MgCl_2$ and lanes 13 and 14 contained 4 mM $MgCl_2$. These results show that both $MnCl_2$ and $MgCl_2$ can be used as the necessary divalent cation to enable the cleavage activity of the Cleavase® A/G enzyme in these reactions and that the invasive cleavage reaction can tolerate a broad range of concentrations of these components.

In addition to examining the effects of the salt environment on the rate of product accumulation in the invasive cleavage reaction, the use of reaction constituents shown to be effective in enhancing nucleic acid hybridization in either standard hybridization assays (*e.g.*, blot hybridization) or in ligation reactions was examined. These components may act as volume excluders, increasing the effective concentration of the nucleic acids of interest and thereby enhancing hybridization, or they may act as charge-shielding agents to minimize repulsion between the highly charged backbones of the nucleic acids strands. The results of these experiments are described in Sections g and h below.

g) Effect Of CTAB Addition

The polycationic detergent cetyltrietheylammonium bromide (CTAB) has been shown to dramatically enhance hybridization of nucleic acids (Pontius and Berg, Proc. Natl. Acad. Sci. USA 88:8237 [1991]). The effect of adding the detergent CTAB in concentrations from 100 mM to 1 mM to invasive cleavage reactions in which 150 mM LiCl was used in place of the KCl in otherwise standard reactions was also investigated. These results showed that 200 mM CTAB may have a very moderate enhancing effect under these reaction conditions, and the presence of CTAB in excess of about 500 μM was inhibitory to the accumulation of specific cleavage product.

h) Effect Of PEG Addition

The effect of adding polyethylene glycol (PEG) at 4.8 or 12% (w/v) concentrations to otherwise standard reactions was also examined. The effects of increasing the reaction temperature of the PEG-containing reactions was examined by performing duplicate sets of PEG titration reactions at 61°C and 65°C. The results showed that at all percentages tested, and at both temperatures tested, the inclusion of PEG substantially eliminated the production of specific cleavage product.

In addition to, the presence of 1X Denhardts in the reaction mixture was found to have no adverse effect upon the cleavage reaction (50X Denhardts contains per 500 ml: 5 g Ficoll, 5 g polyvinylpyrrolidone, 5 g BSA). Further , the presence of each component of Denhardt's was examined individually (*i.e.*, Ficoll alone, polyvinylpyrrolidone alone, BSA alone) for the effect upon the Invader™-directed cleavage reaction; no adverse effect was observed.

i) Effect Of The Addition Of Stabilizing Agents

Another approach to enhancing the output of the invasive cleavage reaction is to enhance the activity of the enzyme employed, either by increasing its stability in the reaction environment or by increasing its turnover rate. Without regard to the precise mechanism by which various agents operate in the invasive cleavage reaction, a number of agents commonly used to stabilize enzymes during prolonged storage were tested for the ability to enhance the accumulation of specific cleavage product in the invasive cleavage reaction.

The effects of adding glycerol at 15% and of adding the detergents Tween-20 and Nonidet-P40 at 1.5%, alone or in combination, in otherwise standard reactions were also examined. The results demonstrated that under these conditions these adducts had little or no effect on the accumulation of specific cleavage product.

The effects of adding gelatin to reactions in which the salt identity and concentration were varied from the standard reaction were also investigated. The results demonstrated that in the absence of salt the gelatin had a moderately enhancing effect on the accumulation of specific cleavage product, but when either salt (KCl or LiCl) was added to reactions performed under these conditions, increasing amounts of gelatin reduced the product accumulation.

j) Effect Of Adding Large Amounts Of Non-Target Nucleic Acid

In detecting specific nucleic acid sequences within samples, it is important to determine if the presence of additional genetic material (*i.e.*, non-target nucleic acids) will have a negative effect on the specificity of the assay. In this experiment, the effect of including large amounts of non-target nucleic acid, either DNA or RNA, on the specificity of the invasive cleavage reaction was examined. The data was examined for either an alteration in the expected site of cleavage, or for an increase in the nonspecific degradation of the probe oligonucleotide.

Fig. 40 shows the effects of adding non-target nucleic acid (*e.g.*, genomic DNA or tRNA) to an invasive cleavage reaction performed at 65°C, with 150 mM LiCl in place of the KCl in the standard reaction. The reactions assayed in lanes 1 and 2 contained 235 and 470 ng of genomic DNA, respectively. The reactions analyzed in lanes 3, 4, 5 and 6 contained 100 ng, 200 ng, 500 ng and 1 μg of tRNA, respectively. Lane 7 represents a control reaction which contained no added nucleic acid beyond the amounts used in the standard reaction. The results shown in Fig. 40 demonstrate that the inclusion of non-target nucleic acid in large amounts could visibly slow the accumulation of specific cleavage product (while not limiting the invention to any particular mechanism, it is thought that the additional nucleic acid competes for binding of the enzyme with the specific reaction components). In additional experiments it was found that the effect of adding large amounts of non-target nucleic acid can be compensated for by increasing the enzyme in the reaction. The data shown in Fig. 40 also demonstrate that a key feature of the invasive cleavage reaction, the specificity of the detection, was not compromised by the presence of large amounts of non-target nucleic acid.

In addition to the data presented above, invasive cleavage reactions were run with succinate buffer at pH 5.9 in place of the MOPS buffer used in the "standard" reaction; no adverse effects were observed.

The data shown in Figs. 38-40 and described above demonstrate that the invasive cleavage reaction can be performed using a wide variety of reaction conditions and is therefore suitable for practice in clinical laboratories.

EXAMPLE 19

Detection Of RNA Targets By Invader™-Directed Cleavage

In addition to the clinical need to detect specific DNA sequences for infectious and genetic diseases, there is a need for technologies that can quantitatively detect target nucleic acids that are composed of RNA. For example, a number of viral agents, such as hepatitis C virus (HCV) and human immunodeficiency virus (HIV) have RNA genomic material, the quantitative detection of which can be used as a measure of viral load in a patient sample. Such information can be of critical diagnostic or prognostic value.

Hepatitis C virus (HCV) infection is the predominant cause of post-transfusion non-A, non-B (NANB) hepatitis around the world. In addition, HCV is the major etiologic agent of hepatocellular carcinoma (HCC) and chronic liver disease world wide. The genome of HCV is a small (9.4 kb) RNA molecule. In studies of transmission of HCV by blood transfusion it has been found the presence of HCV antibody, as measured in standard immunological tests, does not always correlate with the infectivity of the sample, while the presence of HCV RNA in a blood sample strongly correlates with infectivity. Conversely, serological tests may remain negative in immunosuppressed infected individuals, while HCV RNA may be easily detected (Cuthbert, Clin. Microbiol. Rev., 7:505 [1994]).

The need for and the value of developing a probe-based assay for the detection the HCV RNA is clear. The polymerase chain reaction has been used to detect HCV in clinical samples, but the problems associated with carry-over contamination of samples has been a concern. Direct detection of the viral RNA without the need to perform either reverse transcription or amplification would allow the elimination of several of the points at which existing assays may fail.

The genome of the positive-stranded RNA hepatitis C virus comprises several regions including 5' and 3' noncoding regions (*i.e.*, 5' and 3' untranslated regions) and a polyprotein coding region which encodes the core protein (C), two envelope glycoproteins (E1 and E2/NS1) and six nonstructural glycoproteins (NS2-NS5b). Molecular biological analysis of the HCV genome has showed that some regions of the genome are very highly conserved between isolates, while other regions are fairly rapidly changeable. The 5' noncoding region (NCR) is the most highly conserved region in the HCV. These analyses have allowed these viruses to be divided into six basic genotype groups, and then further classified into over a dozen sub-types (the nomenclature and division of HCV genotypes is evolving; see Altamirano *et al.*, J. Infect. Dis., 171:1034 (1995) for a recent classification scheme).

In order to develop a rapid and accurate method of detecting HCV present in infected individuals, the ability of the Invader™-directed cleavage reaction to detect HCV RNA was examined. Plasmids containing DNA derived from the conserved 5'-untranslated region of six different HCV RNA isolates were used to generate templates for *in vitro* transcription. The HCV sequences contained within these six plasmids represent genotypes 1 (four sub-types represented; 1a, 1b, 1c, and Δ1c), 2, and 3. The nomenclature of the HCV genotypes used herein is that of Simmonds *et al.* (as described in Altamirano *et al., supra*). The Δ1c subtype was used in the model detection reaction described below.

a) Generation Of Plasmids Containing HCV Sequences

Six DNA fragments derived from HCV were generated by RT-PCR using RNA extracted from serum samples of blood donors; these PCR fragments were a gift of Dr. M. Altamirano (University of British Columbia, Vancouver). These PCR fragments represent HCV sequences derived from HCV genotypes 1a, 1b, 1c, Δ1c, 2c and 3a.

The RNA extraction, reverse transcription and PCR were performed using standard techniques (Altamirano *et al., supra*). Briefly, RNA was extracted from 100 μl of serum using guanidine isothiocyanate, sodium lauryl sarkosate and phenol-chloroform (Inchauspe *et al.*, Hepatol., 14:595 [1991]). Reverse transcription was performed according to the manufacturer's instructions using a GeneAmp rTh reverse transcriptase RNA PCR kit (Perkin-Elmer) in the presence of an external antisense primer, HCV342. The sequence of the HCV342 primer is 5'-GGTTTTTCTTTGAGGTTTAG-3' (SEQ ID NO:40). Following termination of the RT reaction, the sense primer HCV7 (5'-GCGACACTCCACCATAGAT-3' [SEQ ID NO:41]) and magnesium were added and a first PCR was performed. Aliquots of the first PCR products were used in a second (nested) PCR in the presence of primers HCV46 (5'-CTGTCTTCACGCAGAAAGC-3' [SEQ ID NO:42]) and HCV308 [5'-GCACGGT CTACGAGACCTC-3' [SEQ ID NO:43]). The PCRs produced a 281 bp product which corresponds to a conserved 5' noncoding region (NCR) region of HCV between positions -284 and -4 of the HCV genome (Altramirano *et al., supra*).

The six 281 bp PCR fragments were used directly for cloning or they were subjected to an additional amplification step using a 50 μl PCR comprising approximately 100 fmoles of DNA, the HCV46 and HCV308 primers at 0.1 μM, 100 μM of all four dNTPs and 2.5 units of *Taq* DNA polymerase in a buffer containing 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$ and 0.1% Tween 20. The PCRs were cycled 25 times at 96°C for 45 sec.,

55°C for 45 sec. and 72°C for 1 min. Two microliters of either the original DNA samples or the reamplified PCR products were used for cloning in the linear pT7Blue T-vector (Novagen) according to manufacturer's protocol. After the PCR products were ligated to the pT7Blue T-vector, the ligation reaction mixture was used to transform competent JM109 cells (Promega). Clones containing the pT7Blue T-vector with an insert were selected by the presence of colonies having a white color on LB plates containing 40 μg/ml X-Gal, 40 μg/ml IPTG and 50 μg/ml ampicillin. Four colonies for each PCR sample were picked and grown overnight in 2 ml LB media containing 50 μg/ml carbenicillin. Plasmid DNA was isolated using the following alkaline miniprep protocol. Cells from 1.5 ml of the overnight culture were collected by centrifugation for 2 min. in a microcentrifuge (14K rpm), the supernatant was discarded and the cell pellet was resuspended in 50 μl TE buffer with 10 μg/ml RNAse A (Pharmacia). One hundred microliters of a solution containing 0.2 N NaOH, 1% SDS was added and the cells were lysed for 2 min. The lysate was gently mixed with 100 μl of 1.32 M potassium acetate, pH 4.8, and the mixture was centrifuged for 4 min. in a microcentrifuge (14K rpm); the pellet comprising cell debris was discarded. Plasmid DNA was precipitated from the supernatant with 200 μl ethanol and pelleted by centrifugation a microcentrifuge (14K rpm). The DNA pellet was air dried for 15 min. and was then redissolved in 50 μl TE buffer (10 mM Tris-HCl, pH 7.8, 1 mM EDTA).

b) Reamplification Of HCV Clones To Add The Phage T7 Promoter For Subsequent *In Vitro* Transcription

To ensure that the RNA product of transcription had a discrete 3' end it was necessary to create linear transcription templates which stopped at the end of the HCV sequence. These fragments were conveniently produced using the PCR to reamplify the segment of the plasmid containing the phage promoter sequence and the HCV insert. For these studies, the clone of HCV type Δ1c was reamplified using a primer that hybridizes to the T7 promoter sequence: 5'-TAATACGACTCACTATAGGG-3' (SEQ ID NO:44; "the T7 promoter primer") (Novagen) in combination with the 3' terminal HCV-specific primer HCV308 (SEQ ID NO:43). For these reactions, 1 μl of plasmid DNA (approximately 10 to 100 ng) was reamplified in a 200 μl PCR using the T7 and HCV308 primers as described above with the exception that 30 cycles of amplification were employed. The resulting amplicon was 354 bp in length. After amplification the PCR mixture was transferred to a fresh 1.5 ml microcentrifuge tube, the mixture was brought to a final concentration of 2 M NH_4OAc, and

the products were precipitated by the addition of one volume of 100% isopropanol. Following a 10 min. incubation at room temperature, the precipitates were collected by centrifugation, washed once with 80% ethanol and dried under vacuum. The collected material was dissolved in 100 μl nuclease-free distilled water (Promega).

Segments of RNA were produced from this amplicon by *in vitro* transcription using the RiboMAX™ Large Scale RNA Production System (Promega) in accordance with the manufacturer's instructions, using 5.3 μg of the amplicon described above in a 100 μl reaction. The transcription reaction was incubated for 3.75 hours, after which the DNA template was destroyed by the addition of 5-6 μl of RQ1 RNAse-free DNAse (1 unit/μl) according to the RiboMAX™ kit instructions. The reaction was extracted twice with phenol/chloroform/isoamyl alcohol (50:48:2) and the aqueous phase was transferred to a fresh microcentrifuge tube. The RNA was then collected by the addition of 10 μl of 3M NH_4OAc, pH 5.2 and 110 μl of 100% isopropanol. Following a 5 min. incubation at 4°C, the precipitate was collected by centrifugation, washed once with 80% ethanol and dried under vacuum. The sequence of the resulting RNA transcript (HCV1.1 transcript) is listed in SEQ ID NO:45.

c) Detection Of The HCV1.1 Transcript In The Invader™-Directed Cleavage Assay

Detection of the HCV1.1 transcript was tested in the Invader™-directed cleavage assay using an HCV-specific probe oligonucleotide (5'-CCGGTCGTCCTGGCAAT XCC-3' [SEQ ID NO:46]); X indicates the presence of a fluorescein dye on an abasic linker) and an HCV-specific Invader™ oligonucleotide (5'-GTTTATCCAAGAAAGGAC CCGGTC-3' [SEQ ID NO:47]) that causes a 6-nucleotide invasive cleavage of the probe.

Each 10 μl of reaction mixture comprised 5 pmole of the probe oligonucleotide (SEQ ID NO:46) and 10 pmole of the Invader™ oligonucleotide (SEQ ID NO:47) in a buffer of 10 mM MOPS, pH 7.5 with 50 mM KCl, 4 mM $MnCl_2$, 0.05% each Tween-20 and Nonidet-P40 and 7.8 units RNasin® ribonuclease inhibitor (Promega). The cleavage agents employed were Cleavase® A/G (used at 5.3 ng/10 μl reaction) or DNAPTth (used at 5 polymerase units/10 μl reaction). The amount of RNA target was varied as indicated below. When RNAse treatment is indicated, the target RNAs were pre-treated with 10 μg of RNase A (Sigma) at 37°C for 30 min. to demonstrate that the detection was specific for the RNA in the reaction

and not due to the presence of any residual DNA template from the transcription reaction. RNase-treated aliquots of the HCV RNA were used directly without intervening purification.

For each reaction, the target RNAs were suspended in the reaction solutions as described above, but lacking the cleavage agent and the $MnCl_2$ for a final volume of 10 μl, with the Invader™ and probe at the concentrations listed above. The reactions were warmed to 46°C and the reactions were started by the addition of a mixture of the appropriate enzyme with $MnCl_2$. After incubation for 30 min. at 46°C, the reactions were stopped by the addition of 8 μl of 95% formamide, 10 mM EDTA and 0.02% methyl violet (methyl violet loading buffer). Samples were then resolved by electrophoresis through a 15% denaturing polyacrylamide gel (19:1 cross-linked), containing 7 M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA. Following electrophoresis, the labeled reaction products were visualized using the FMBIO-100 Image Analyzer (Hitachi), with the resulting imager scan shown in Fig. 41.

In Fig. 41, the samples analyzed in lanes 1-4 contained 1 pmole of the RNA target, the reactions shown in lanes 5-8 contained 100 fmoles of the RNA target and the reactions shown in lanes 9-12 contained 10 fmoles of the RNA target. All odd-numbered lanes depict reactions performed using Cleavase® A/G enzyme and all even-numbered lanes depict reactions performed using DNAPTth. The reactions analyzed in lanes 1, 2, 5, 6, 9 and 10 contained RNA that had been pre-digested with RNase A. These data demonstrate that the invasive cleavage reaction efficiently detects RNA targets and further, the absence of any specific cleavage signal in the RNase-treated samples confirms that the specific cleavage product seen in the other lanes is dependent upon the presence of input RNA.

EXAMPLE 20

The Fate Of The Target RNA In The Invader™-Directed Cleavage Reaction

In this Example, the fate of the RNA target in the Invader™-directed cleavage reaction was examined. As shown above in Example 1D, when RNAs are hybridized to DNA oligonucleotides, the 5' nucleases associated with DNA polymerases can be used to cleave the RNAs; such cleavage can be suppressed when the 5' arm is long or when it is highly structured (Lyamichev *et al.*, Science 260:778 [1993], and U.S. Patent No. 5,422,253, the disclosure of which is herein incorporated by reference). In this experiment, the extent to which the RNA target would be cleaved by the cleavage agents when hybridized to the

detection oligonucleotides (*i.e.*, the probe and Invader™ oligonucleotides) was examined using reactions similar to those described in Example 20, performed using fluorescein-labeled RNA as a target.

Transcription reactions were performed as described in Example 19 with the exception that 2% of the UTP in the reaction was replaced with fluorescein-12-UTP (Boehringer Mannheim) and 5.3 μg of the amplicon was used in a 100 μl reaction. The transcription reaction was incubated for 2.5 hours, after which the DNA template was destroyed by the addition of 5-6 μl of RQ1 RNAse-free DNAse (1 unit/μl) according to the RiboMAX™ kit instructions. The organic extraction was omitted and the RNA was collected by the addition of 10 μl of 3M NaOAc, pH 5.2 and 110 μl of 100% isopropanol. Following a 5 min. incubation at 4°C, the precipitate was collected by centrifugation, washed once with 80% ethanol and dried under vacuum. The resulting RNA was dissolved in 100 μl of nuclease-free water. Half (*i.e.*, 50%) of the sample was purified by electrophoresis through a 8% denaturing polyacrylamide gel (19:1 cross-linked), containing 7 M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA. The gel slice containing the full-length material was excised and the RNA was eluted by soaking the slice overnight at 4°C in 200 μl of 10 mM Tris-Cl, pH 8.0, 0.1 mM EDTA and 0.3 M NaOAc. The RNA was then precipitated by the addition of 2.5 volumes of 100% ethanol. After incubation at -20°C for 30 min., the precipitates were recovered by centrifugation, washed once with 80% ethanol and dried under vacuum. The RNA was dissolved in 25 μl of nuclease-free water and then quantitated by UV absorbance at 260 nm.

Samples of the purified RNA target were incubated for 5 or 30 min. in reactions that duplicated the Cleavase® A/G and DNAPTth Invader™ reactions described in Example 20 with the exception that the reactions lacked probe and Invader™ oligonucleotides. Subsequent analysis of the products showed that the RNA was very stable, with a very slight background of non-specific degradation, appearing as a gray background in the gel lane. The background was not dependent on the presence of enzyme in the reaction.

Invader™ detection reactions using the purified RNA target were performed using the probe/Invader™ pair described in Example 19 (SEQ ID NOS:46 and 47). Each reaction included 500 fmole of the target RNA, 5 pmoles of the fluorescein-labeled probe and 10 pmoles of the Invader™ oligonucleotide in a buffer of 10 mM MOPS, pH 7.5 with 150 mM LiCl, 4 mM $MnCl_2$, 0.05% each Tween-20 and Nonidet-P40 and 39 units RNAsin® (Promega). These components were combined and warmed to 50°C and the reactions were

started by the addition of either 53 ng of Cleavase® A/G or 5 polymerase units of DNAPTth. The final reaction volume was 10 μl. After 5 min at 50°C, 5 μl aliquots of each reaction were removed to tubes containing 4 μl of 95% formamide, 10 mM EDTA and 0.02% methyl violet. The remaining aliquot received a drop of ChillOut® evaporation barrier and was incubated for an additional 25 min. These reactions were then stopped by the addition of 4 μl of the above formamide solution. The products of these reactions were resolved by electrophoresis through separate 20% denaturing polyacrylamide gels (19:1 cross-linked), containing 7 M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA. Following electrophoresis, the labeled reaction products were visualized using the FMBIO-100 Image Analyzer (Hitachi), with the resulting imager scans shown in Figs. 42A (5 min reactions) and 42B (30 min. reactions).

In Fig. 53 the target RNA is seen very near the top of each lane, while the labeled probe and its cleavage products are seen just below the middle of each panel. The FMBIO-100 Image Analyzer was used to quantitate the fluorescence signal in the probe bands. In each panel, lane 1 contains products from reactions performed in the absence of a cleavage agent, lane 2 contains products from reactions performed using Cleavase® A/G and lane 3 contains products from reactions performed using DNAPTth.

Quantitation of the fluorescence signal in the probe bands revealed that after a 5 min. incubation, 12% or 300 fmole of the probe was cleaved by the Cleavase® A/G and 29% or 700 fmole was cleaved by the DNAPTth. After a 30 min. incubation, Cleavase® A/G had cleaved 32% of the probe molecules and DNAPTth had cleaved 70% of the probe molecules. (The images shown in Figs. 42A and 42B were printed with the intensity adjusted to show the small amount of background from the RNA degradation, so the bands containing strong signals are saturated and therefore these images do not accurately reflect the differences in measured fluorescence)

The data shown in Fig. 42 clearly shows that, under invasive cleavage conditions, RNA molecules are sufficiently stable to be detected as a target and that each RNA molecule can support many rounds of probe cleavage.

EXAMPLE 21

Titration Of Target RNA In The Invader™-Directed Cleavage Assay

One of the primary benefits of the Invader™-directed cleavage assay as a means for detection of the presence of specific target nucleic acids is the correlation between the amount of cleavage product generated in a set amount of time and the quantity of the nucleic acid of interest present in the reaction. The benefits of quantitative detection of RNA sequences was discussed in Example 19. In this Example, the quantitative nature of the detection assay was demonstrated through the use of various amounts of target starting material. In addition to demonstrating the correlation between the amounts of input target and output cleavage product, these data graphically show the degree to which the RNA target can be recycled in this assay

The RNA target used in these reactions was the fluorescein-labeled material described in Example 20 (*i.e.*, SEQ ID NO:45). Because the efficiency of incorporation of the fluorescein-12-UTP by the T7 RNA polymerase was not known, the concentration of the RNA was determined by measurement of absorbance at 260 nm, not by fluorescence intensity. Each reaction comprised 5 pmoles of the fluorescein-labeled probe (SEQ ID NO:46) and 10 pmoles of the Invader™ oligonucleotide (SEQ ID NO:47) in a buffer of 10 mM MOPS, pH 7.5 with 150 mM LiCl, 4 mM $MnCl_2$, 0.05% each Tween-20 and Nonidet-P40 and 39 units of RNAsin® (Promega). The amount of target RNA was varied from 1 to 100 fmoles, as indicated below. These components were combined, overlaid with ChillOut® evaporation barrier and warmed to 50°C; the reactions were started by the addition of either 53 ng of Cleavase® A/G or 5 polymerase units of DNAPTth, to a final reaction volume of 10 µl. After 30 minutes at 50°C, reactions were stopped by the addition of 8 µl of 95% formamide, 10 mM EDTA and 0.02% methyl violet. The unreacted markers in lanes 1 and 2 were diluted in the same total volume (18 µl). The samples were heated to 90°C for 1 minute and 2.5 µl of each of these reactions were resolved by electrophoresis through a 20% denaturing polyacrylamide gel (19:1 cross link) with 7M urea in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA, and the labeled reaction products were visualized using the FMBIO-100 Image Analyzer (Hitachi), with the resulting imager scans shown in Fig. 43.

In Fig. 43, lanes 1 and 2 show 5 pmoles of uncut probe and 500 fmoles of untreated RNA, respectively. The probe is the very dark signal near the middle of the panel, while the RNA is the thin line near the top of the panel. These RNAs were transcribed with a 2%

substitution of fluorescein-12-UTP for natural UTP in the transcription reaction. The resulting transcript contains 74 U residues, which would give an average of 1.5 fluorescein labels per molecule. With one tenth the molar amount of RNA loaded in lane 2, the signal in lane 2 should be approximately one seventh (0.15X) the fluorescence intensity of the probe in lane 1. Measurements indicated that the intensity was closer to one fortieth, indicating an efficiency of label incorporation of approximately 17%. Because the RNA concentration was verified by A260 measurement this does not alter the experimental observations below, but it should be noted that the signal from the RNA and the probes does not accurately reflect the relative amounts in the reactions.

The reactions analyzed in lanes 3 through 7 contained 1, 5, 10, 50 and 100 fmoles of target, respectively, with cleavage of the probe accomplished by Cleavase® A/G. The reactions analyzed in lanes 8 through 12 repeated the same array of target amounts, with cleavage of the probe accomplished by DNAPTth. The boxes seen surrounding the product bands show the area of the scan in which the fluorescence was measured for each reaction. The number of fluorescence units detected within each box is indicated below each box; background florescence was also measured.

It can be seen by comparing the detected fluorescence in each lane that the amount of product formed in these 30 minute reactions can be correlated to the amount of target material. The accumulation of product under these conditions is slightly enhanced when DNAPTth is used as the cleavage agent, but the correlation with the amount of target present remains. This demonstrates that the Invader™ assay can be used as a means of measuring the amount of target RNA within a sample.

Comparison of the fluorescence intensity of the input RNA with that of the cleaved product shows that the Invader™-directed cleavage assay creates signal in excess of the amount of target, so that the signal visible as cleaved probe is far more intense than that representing the target RNA. This further confirms the results described in Example 20, in which it was demonstrated that each RNA molecule could be used many times.

EXAMPLE 22

Detection Of DNA By Charge Reversal

The detection of specific targets is achieved in the Invader™-directed cleavage assay by the cleavage of the probe oligonucleotide. In addition to the methods described in the preceding Examples, the cleaved probe may be separated from the uncleaved probe using the charge reversal technique described below. This novel separation technique is related to the observation that positively charged adducts can affect the electrophoretic behavior of small oligonucleotides because the charge of the adduct is significant relative to charge of the whole complex. Observations of aberrant mobility due to charged adducts have been reported in the literature, but in all cases found, the applications pursued by other scientists have involved making oligonucleotides larger by enzymatic extension. As the negatively charged nucleotides are added on, the positive influence of the adduct is reduced to insignificance. As a result, the effects of positively charged adducts have been dismissed and have received infinitesimal notice in the existing literature.

This observed effect is of particular utility in assays based on the cleavage of DNA molecules. When an oligonucleotide is shortened through the action of a Cleavase® enzyme or other cleavage agent, the positive charge can be made to not only significantly reduce the net negative charge, but to actually override it, effectively "flipping" the net charge of the labeled entity. This reversal of charge allows the products of target-specific cleavage to be partitioned from uncleaved probe by extremely simple means. For example, the products of cleavage can be made to migrate towards a negative electrode placed at any point in a reaction vessel, for focused detection without gel-based electrophoresis. When a slab gel is used, sample wells can be positioned in the center of the gel, so that the cleaved and uncleaved probes can be observed to migrate in opposite directions. Alternatively, a traditional vertical gel can be used, but with the electrodes reversed relative to usual DNA gels (*i.e.*, the positive electrode at the top and the negative electrode at the bottom) so that the cleaved molecules enter the gel, while the uncleaved disperse into the upper reservoir of electrophoresis buffer.

An additional benefit of this type of readout is that the absolute nature of the partition of products from substrates means that an abundance of uncleaved probe can be supplied to drive the hybridization step of the probe-based assay, yet the unconsumed probe can be subtracted from the result to reduce background.

Through the use of multiple positively charged adducts, synthetic molecules can be constructed with sufficient modification that the normally negatively charged strand is made nearly neutral. When so constructed, the presence or absence of a single phosphate group can mean the difference between a net negative or a net positive charge. This observation has particular utility when one objective is to discriminate between enzymatically generated fragments of DNA, which lack a 3' phosphate, and the products of thermal degradation, which retain a 3' phosphate (and thus two additional negative charges).

a) Characterization Of The Products Of Thermal Breakage Of DNA Oligonucleotides

Thermal degradation of DNA probes results in high background which can obscure signals generated by specific enzymatic cleavage, decreasing the signal-to-noise ratio. To better understand the nature of DNA thermal degradation products, the 5' tetrachloro-fluorescein (TET)-labeled oligonucleotides 78 (SEQ ID NO:48) and 79 (SEQ ID NO:49) (100 pmole each) were incubated in 50 µl 10 mM $NaCO_3$ (pH 10.6), 50 mM NaCl at 90°C for 4 hours. To prevent evaporation of the samples, the reaction mixture was overlaid with 50 µl of ChillOut® liquid wax. The reactions were then divided in two equal aliquots (A and B). Aliquot A was mixed with 25 µl of methyl violet loading buffer and Aliquot B was dephosphorylated by addition of 2.5 µl of 100 mM $MgCl_2$ and 1 µl of 1 unit/µl Calf Intestinal Alkaline Phosphatase (CIAP) (Promega), with incubation at 37°C for 30 min. after which 25 µl of methyl violet loading buffer was added. One microliter of each sample was resolved by electrophoresis through a 12% polyacrylamide denaturing gel and imaged as described in Example 21; a 585 nm filter was used with the FMBIO Image Analyzer. The resulting imager scan is shown in Fig. 44.

In Fig. 44, lanes 1-3 contain the TET-labeled oligonucleotide 78 and lanes 4-6 contain the TET-labeled oligonucleotides 79. Lanes 1 and 4 contain products of reactions which were not heat treated. Lanes 2 and 5 contain products from reactions which were heat treated and lanes 3 and 6 contain products from reactions which were heat treated and subjected to phosphatase treatment.

As shown in Fig. 44, heat treatment causes significant breakdown of the 5'-TET-labeled DNA, generating a ladder of degradation products (Fig. 44, lanes 2, 3, 5 and 6). Band intensities correlate with purine and pyrimidine base positioning in the oligonucleotide sequences, indicating that backbone hydrolysis may occur through formation of abasic

intermediate products that have faster rates for purines than for pyrimidines (Lindahl and Karlström, Biochem., 12:5151 [1973]).

Dephosphorylation decreases the mobility of all products generated by the thermal degradation process, with the most pronounced effect observed for the shorter products (Fig. 44, lanes 3 and 6). This demonstrates that thermally degraded products possess a 3' end terminal phosphoryl group which can be removed by dephosphorylation with CIAP. Removal of the phosphoryl group decreases the overall negative charge by 2. Therefore, shorter products which have a small number of negative charges are influenced to a greater degree upon the removal of two charges. This leads to a larger mobility shift in the shorter products than that observed for the larger species.

The fact that the majority of thermally degraded DNA products contain 3' end phosphate groups and Cleavase® enzyme-generated products do not allowed the development of simple isolation methods for products generated in the Invader™-directed cleavage assay. The extra two charges found in thermal breakdown products do not exist in the specific cleavage products. Therefore, if one designs assays that produce specific products which contain a net positive charge of one or two, then similar thermal breakdown products will either be negative or neutral. The difference can be used to isolate specific products by reverse charge methods as shown below.

b) Dephosphorylation Of Short Amino-Modified Oligonucleotides Can Reverse The Net Charge Of The Labeled Product

To demonstrate how oligonucleotides can be transformed from net negative to net positively charged compounds, the four short amino-modified oligonucleotides labeled **70**, **74**, **75** and **76** and shown in Figs. 45-47 were synthesized (Fig. 45 shows both oligonucleotides **70** and **74**). All four modified oligonucleotides possess Cy-3 dyes positioned at the 5'-end which individually are positively charged under reaction and isolation conditions described in this Example. Compounds **70** and **74** contain two amino modified thymidines that, under reaction conditions, display positively charged $R\text{-}NH_3^+$ groups attached at the C5 position through a C_{10} or C_6 linker, respectively. Because compounds **70** and **74** are 3'-end phosphorylated, they consist of four negative charges and three positive charges. Compound **75** differs from **74** in that the internal C_6 amino modified thymidine phosphate in **74** is replaced by a thymidine methyl phosphonate. The phosphonate backbone is uncharged and so

there are a total of three negative charges on compound **75**. This gives compound **75** a net negative one charge. Compound **76** differs from **70** in that the internal amino modified thymidine is replaced by an internal cytosine phosphonate. The pK_a of the N3 nitrogen of cytosine can be from 4 to 7. Thus, the net charges of this compound, can be from -1 to 0 depending on the pH of the solution. For the simplicity of analysis, each group is assigned a whole number of charges, although it is realized that, depending on the pK_a of each chemical group and ambient pH, a real charge may differ from the whole number assigned. It is assumed that this difference is not significant over the range of pHs used in the enzymatic reactions studied here.

Dephosphorylation of these compounds, or the removal of the 3' end terminal phosphoryl group, results in elimination of two negative charges and generates products that have a net positive charge of one. In this experiment, the method of isoelectric focusing (IEF) was used to demonstrate a change from one negative to one positive net charge for the described substrates during dephosphorylation.

Substrates **70**, **74**, **75** and **76** were synthesized by standard phosphoramidite chemistries and deprotected for 24 hours at 22°C in 14 M aqueous ammonium hydroxide solution, after which the solvent was removed in vacuo. The dried powders were resuspended in 200 µl of H_2O and filtered through 0.2 µm filters. The concentration of the stock solutions was estimated by UV-absorbance at 261 nm of samples diluted 200-fold in H_2O using a spectrophotometer (Spectronic Genesys 2, Milton Roy, Rochester, NY).

Dephosphorylation of compounds **70** and **74**, **75** and **76** was accomplished by treating 10 µl of the crude stock solutions (ranging in concentration from approximately 0.5 to 2 mM) with 2 units of CIAP in 100 µl of CIAP buffer (Promega) at 37°C for 1 hour. The reactions were then heated to 75°C for 15 min. in order to inactivate the CIAP. For clarity, dephosphorylated compounds are designated 'dp'. For example, after dephosphorylation, substrate **70** becomes **70dp**.

To prepare samples for IEF experiments, the concentration of the stock solutions of substrate and dephosphorylated product were adjusted to a uniform absorbance of 8.5×10^{-3} at 532 nm by dilution with water. Two microliters of each sample were analyzed by IEF using a PhastSystem electrophoresis unit (Pharmacia) and PhastGel IEF 3-9 media (Pharmacia) according to the manufacturer's protocol. Separation was performed at 15°C with the following program: pre-run; 2,000 V, 2.5 mA, 3.5 W, 75 Vh; load; 200 V, 2.5 mA, 3.5 W, 15 Vh; run; 2,000 V; 2.5 mA; 3.5 W, 130 Vh. After separation, samples were visualized by

using the FMBIO Image Analyzer (Hitachi) fitted with a 585 nm filter. The resulting imager scan is shown in Fig. 48.

Fig. 48 shows results of IEF separation of substrates **70**, **74**, **75** and **76** and their dephosphorylated products. The arrow labeled "Sample Loading Position" indicates a loading line, the '+' sign shows the position of the positive electrode and the '-' sign indicates the position of the negative electrode.

The results shown in Fig. 48 demonstrate that substrates **70**, **74**, **75** and **76** migrated toward the positive electrode, while the dephosphorylated products **70dp**, **74dp**, **75dp** and **76dp** migrated toward negative electrode. The observed differences in mobility direction was in accord with predicted net charge of the substrates (minus one) and the products (plus one). Small perturbations in the mobilities of the phosphorylated compounds indicate that the overall pIs vary. This was also true for the dephosphorylated compounds. The presence of the cytosine in **76dp**, for instance, moved this compound further toward the negative electrode which was indicative of a higher overall pI relative to the other dephosphorylated compounds. It is important to note that additional positive charges can be obtained by using a combination of natural amino modified bases (**70dp** and **74dp**) along with uncharged methylphosphonate bridges (products **75dp** and **76dp**).

The results shown above demonstrate that the removal of a single phosphate group can flip the net charge of an oligonucleotide to cause reversal in an electric field, allowing easy separation of products, and that the precise base composition of the oligonucleotides affect absolute mobility but not the charge-flipping effect.

EXAMPLE 23

Detection Of Specific Cleavage Products In The Invader™-Directed Cleavage Reaction By Charge Reversal

In this Example the ability to isolate products generated in the Invader™-directed cleavage assay from all other nucleic acids present in the reaction cocktail was demonstrated using charge reversal. This experiment utilized the following Cy3-labeled oligonucleotide: 5'-Cy3-AminoT-AminoT-CTTTTCACCAGCGAGACGGG-3' (SEQ ID NO:50; termed "oligo 61"). Oligo 61 was designed to release upon cleavage a net positively charged labeled product. To test whether or not a net positively charged 5'-end labeled product would be recognized by the Cleavase® enzymes in the Invader™-directed cleavage assay format, probe oligo 61 (SEQ ID NO:50) and invading oligonucleotide 67 (SEQ ID NO:51) were chemically

synthesized on a DNA synthesizer (ABI 391) using standard phosphoramidite chemistries and reagents obtained from Glen Research (Sterling, VA).

Each assay reaction comprised 100 fmoles of M13mp18 single stranded DNA, 10 pmoles each of the probe (SEQ ID NO:50) and Invader™ (SEQ ID NO:51) oligonucleotides, and 20 units of Cleavase® A/G in a 10 µl solution of 10 mM MOPS, pH 7.4 with 100 mM KCl. Samples were overlaid with mineral oil to prevent evaporation. The samples were brought to either 50°C, 55°C, 60°C, or 65°C and cleavage was initiated by the addition of 1 µl of 40 mM $MnCl_2$. Reactions were allowed to proceed for 25 minutes and then were terminated by the addition of 10 µl of 95% formamide containing 20 mM EDTA and 0.02% methyl violet. The negative control experiment lacked the target M13mp18 and was run at 60°C. Five microliters of each reaction were loaded into separate wells of a 20% denaturing polyacrylamide gel (cross-linked 29:1) with 8 M urea in a buffer containing 45 mM Tris-Borate (pH 8.3) and 1.4 mM EDTA. An electric field of 20 watts was applied for 30 minutes, with the electrodes oriented as indicated in Fig. 49B (*i.e.*, in reverse orientation). The products of these reactions were visualized using the FMBIO fluorescence imager and the resulting imager scan is shown in Fig. 49B.

Fig. 49A provides a schematic illustration showing an alignment of the Invader™ (SEQ ID NO:50) and probe (SEQ ID NO:51) along the target M13mp18 DNA; only 53 bases of the M13mp18 sequence is shown (SEQ ID NO:52). The sequence of the Invader™ oligonucleotide is displayed under the M13mp18 target and an arrow is used above the M13mp18 sequence to indicate the position of the Invader™ relative to the probe and target. As shown in Fig. 49A, the Invader™ and probe oligonucleotides share a 2 base region of overlap.

In Fig. 49B, lanes 1-6 contain reactions performed at 50°C, 55°C, 60°C, and 65°C, respectively; lane 5 contained the control reaction (lacking target). In Fig. 49B, the products of cleavage are seen as dark bands in the upper half of the panel; the faint lower band seen appears in proportion to the amount of primary product produced and, while not limiting the invention to a particular mechanism, may represent cleavage one nucleotide into the duplex. The uncleaved probe does not enter the gel and is thus not visible. The control lane showed no detectable signal over background (lane 5). As expected in an invasive cleavage reaction, the rate of accumulation of specific cleavage product was temperature-dependent. Using these particular oligonucleotides and target, the fastest rate of accumulation of product was observed at 55°C (lane 2) and very little product observed at 65°C (lane 4).

When incubated for extended periods at high temperature, DNA probes can break non-specifically (*i.e.*, suffer thermal degradation) and the resulting fragments contribute an interfering background to the analysis. The products of such thermal breakdown are distributed from single-nucleotides up to the full length probe. In this experiment, the ability of charge based separation of cleavage products (*i.e.*, charge reversal) would allow the sensitive separation of the specific products of target-dependent cleavage from probe fragments generated by thermal degradation was examined.

To test the sensitivity limit of this detection method, the target M13mp18 DNA was serially diluted ten fold over than range of 1 fmole to 1 amole. The Invader™ and probe oligonucleotides were those described above (*i.e.*, SEQ ID NOS:50 and 51). The invasive cleavage reactions were run as described above with the following modifications: the reactions were performed at 55°C, 250 mM or 100 mM KGlu was used in place of the 100 mM KCl and only 1 pmole of the Invader™ oligonucleotide was added. The reactions were initiated as described above and allowed to progress for 12.5 hours. A negative control reaction which lacked added M13m18 target DNA was also run. The reactions were terminated by the addition of 10 μl of 95% formamide containing 20 mM EDTA and 0.02% methyl violet, and 5 μl of these mixtures were electrophoresed and visualized as described above. The resulting imager scan is shown in Fig. 50.

In Fig. 50, lane 1 contains the negative control; lanes 2-5 contain reactions performed using 100 mM KGlu; lanes 6-9 contain reactions performed using 250 mM KGlu. The reactions resolved in lanes 2 and 6 contained 1 fmole of target DNA; those in lanes 3 and 7 contained 100 amole of target; those in lanes 4 and 8 contained 10 amole of target and those in lanes 5 and 9 contained 1 amole of target. The results shown in Fig. 50 demonstrate that the detection limit using charge reversal to detect the production of specific cleavage products in an invasive cleavage reaction is at or below 1 attomole or approximately 6.02×10^5 target molecules. No detectable signal was observed in the control lane, which indicates that non-specific hydrolysis or other breakdown products do not migrate in the same direction as enzyme-specific cleavage products. The excitation and emission maxima for Cy3 are 554 and 568, respectively, while the FMBIO Imager Analyzer excites at 532 and detects at 585. Therefore, the limit of detection of specific cleavage products can be improved by the use of more closely matched excitation source and detection filters.

EXAMPLE 24

Devices And Methods For The Separation And Detection Of Charged Reaction Products

This Example is directed at methods and devices for isolating and concentrating specific reaction products produced by enzymatic reactions conducted in solution whereby the reactions generate charged products from either a charge neutral substrate or a substrate bearing the opposite charge borne by the specific reaction product. The methods and devices of this Example allow isolation of, for example, the products generated by the Invader™-directed cleavage assay of the present invention.

The methods and devices of this Example are based on the principle that when an electric field is applied to a solution of charged molecules, the migration of the molecules toward the electrode of the opposite charge occurs very rapidly. If a matrix or other inhibitory material is introduced between the charged molecules and the electrode of opposite charge such that this rapid migration is dramatically slowed, the first molecules to reach the matrix will be nearly stopped, thus allowing the lagging molecules to catch up. In this way a dispersed population of charged molecules in solution can be effectively concentrated into a smaller volume. By tagging the molecules with a detectable moiety (*e.g.*, a fluorescent dye), detection is facilitated by both the concentration and the localization of the analytes. This Example illustrates two embodiments of devices contemplated by the present invention; of course, variations of these devices will be apparent to those skilled in the art and are within the spirit and scope of the present invention.

Fig. 51 depicts one embodiment of a device for concentrating the positively-charged products generated using the methods of the present invention. As shown in Fig. 51, the device comprises a reaction tube (10) which contains the reaction solution (11). One end of each of two thin capillaries (or other tubes with a hollow core) (13A and 13B) are submerged in the reaction solution (11). The capillaries (13A and 13B) may be suspended in the reaction solution (11) such that they are not in contact with the reaction tube itself; one appropriate method of suspending the capillaries is to hold them in place with clamps (not shown). Alternatively, the capillaries may be suspended in the reaction solution (11) such that they are in contact with the reaction tube itself. Suitable capillaries include glass capillary tubes commonly available from scientific supply companies (*e.g.*, Fisher Scientific or VWR Scientific) or from medical supply houses that carry materials for blood drawing and analysis. Though the present invention is not limited to capillaries of any particular inner

diameter, tubes with inner diameters of up to about 1/8 inch (approximately 3 mm) are particularly preferred for use with the present invention; for example, Kimble No. 73811-99 tubes (VWR Scientific) have an inner diameter of 1.1 mm and are a suitable type of capillary tube. Although the capillaries of the device are commonly composed of glass, any nonconductive tubular material, either rigid or flexible, that can contain either a conductive material or a trapping material is suitable for use in the present invention. One example of a suitable flexible tube is Tygon® clear plastic tubing (Part No. R3603; inner diameter = 1/16 inch; outer diameter = 1/8 inch).

As illustrated in Fig. 51, capillary 13A is connected to the positive electrode of a power supply (20) (*e.g.*, a controllable power supply available through the laboratory suppliers listed above or through electronics supply houses like Radio Shack) and capillary 13B is connected to the negative electrode of the power supply (20). Capillary 13B is filled with a trapping material (14) capable of trapping the positively-charged reaction products by allowing minimal migration of products that have entered the trapping material (14). Suitable trapping materials include, but are not limited to, high percentage (*e.g.*, about 20%) acrylamide polymerized in a high salt buffer (0.5 M or higher sodium acetate or similar salt); such a high percentage polyacrylamide matrix dramatically slows the migration of the positively-charged reaction products. Alternatively, the trapping material may comprise a solid, negatively-charged matrix, such as negatively-charged latex beads, that can bind the incoming positively-charged products. It should be noted that any amount of trapping material (14) capable of inhibiting any concentrating the positively-charged reaction products may be used. Thus, while the capillary 13B in Fig. 51 only contains trapping material in the lower, submerged portion of the tube, the trapping material (14) can be present in the entire capillary (13B); similarly, less trapping material (14) could be present than that shown in Fig. 51 because the positively-charged reaction products generally accumulate within a very small portion of the bottom of the capillary (13B). The amount of trapping material need only be sufficient to make contact with the reaction solution (11) and have the capacity to collect the reaction products. When capillary 13B is not completely filled with the trapping material, the remaining space is filled with any conductive material (15); suitable conductive materials are discussed below.

By comparison, the capillary (13A) connected to the positive electrode of the power supply 20 may be filled with any conductive material (15; indicated by the hatched lines in Fig. 51). This may be the sample reaction buffer (*e.g.*, 10 mM MOPS, pH 7.5 with 150 mM

LiCl, 4 mM $MnCl_2$), a standard electrophoresis buffer (*e.g.*, 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA), or the reaction solution (11) itself. The conductive material (15) is frequently a liquid, but a semi-solid material (*e.g.*, a gel) or other suitable material might be easier to use and is within the scope of the present invention. Moreover, that trapping material used in the other capillary (*i.e.*, capillary 13B) may also be used as the conductive material. Conversely, it should be noted that the same conductive material used in the capillary (13A) attached to the positive electrode may also be used in capillary 13B to fill the space above the region containing the trapping material (14) (*see* Fig. 51).

The top end of each of the capillaries (13A and 13B) is connected to the appropriate electrode of the power supply (20) by electrode wire (18) or other suitable material. Fine platinum wire (*e.g.*, 0.1 to 0.4 mm, Aesar Johnson Matthey, Ward Hill, MA) is commonly used as conductive wire because it does not corrode under electrophoresis conditions. The electrode wire (18) can be attached to the capillaries (13A and 13B) by a nonconductive adhesive (not shown), such as the silicone adhesives that are commonly sold in hardware stores for sealing plumbing fixtures. If the capillaries are constructed of a flexible material, the electrode wire (18) can be secured with a small hose clamp or constricting wire (not shown) to compress the opening of the capillaries around the electrode wire. If the conducting material (15) is a gel, an electrode wire (18) can be embedded directly in the gel within the capillary.

The cleavage reaction is assembled in the reaction tube (10) and allowed to proceed therein as described in proceeding Examples (*e.g.*, Examples 22-23). Though not limited to any particular volume of reaction solution (11), a preferred volume is less than 10 ml and more preferably less than 0.1 ml. The volume need only be sufficient to permit contact with both capillaries. After the cleavage reaction is completed, an electric field is applied to the capillaries by turning on the power source (20). As a result, the positively-charged products generated in the course of the Invader™-directed cleavage reaction which employs an oligonucleotide, which when cleaved, generates a positively charged fragment (described in Ex. 23) but when uncleaved bears a net negative charge, migrate to the negative capillary, where their migration is slowed or stopped by the trapping material (14), and the negatively-charged uncut and thermally degraded probe molecules migrate toward the positive electrode. Through the use of this or a similar device, the positively-charged products of the invasive cleavage reaction are separated from the other material (*i.e.*, uncut and thermally degraded probe) and concentrated from a large volume. Concentration of the product in a small

amount of trapping material (14) allows for simplicity of detection, with a much higher signal-to-noise ratio than possible with detection in the original reaction volume. Because the concentrated product is labelled with a detectable moiety like a fluorescent dye, a commercially-available fluorescent plate reader (not shown) can be used to ascertain the amount of product. Suitable plate readers include both top and bottom laser readers. Capillary 13B can be positioned with the reaction tube (10) at any desired position so as to accommodate use with either a top or a bottom plate reading device.

In the alternative embodiment of the present invention depicted in Fig. 52, the procedure described above is accomplished by utilizing only a single capillary (13B). The capillary (13B) contains the trapping material (14) described above and is connected to an electrode wire (18), which in turn is attached to the negative electrode of a power supply (20). The reaction tube (10) has an electrode (25) embedded into its surface such that one surface of the electrode is exposed to the interior of the reaction tube (10) and another surface is exposed to the exterior of the reaction tube. The surface of the electrode (25) on the exterior of the reaction tube is in contact with a conductive surface (26) connected to the positive electrode of the power supply (20) through an electrode wire (18). Variations of the arrangement depicted in Fig. 52 are also contemplated by the present invention. For example, the electrode (25) may be in contact with the reaction solution (11) through the use of a small hole in the reaction tube (10); furthermore, the electrode wire (18) can be directly attached to the electrode wire (18), thereby eliminating the conductive surface (26).

As indicated in Fig. 52, the electrode (25) is embedded in the bottom of a reaction tube (10) such that one or more reaction tubes may be set on the conductive surface (26). This conductive surface could serve as a negative electrode for multiple reaction tubes; such a surface with appropriate contacts could be applied through the use of metal foils (*e.g.*, copper or platinum, Aesar Johnson Matthey, Ward Hill, MA) in much the same way contacts are applied to circuit boards. Because such a surface contact would not be exposed to the reaction sample directly, less expensive metals, such as the copper could be used to make the electrical connections.

The above devices and methods are not limited to separation and concentration of positively charged oligonucleotides. As will be apparent to those skilled in the art, negatively charged reaction products may be separated from neutral or positively charged reactants using the above device and methods with the exception that capillary 13B is attached to the positive

electrode of the power supply (20) and capillary 13A or alternatively, electrode 25, is attached to the negative electrode of the power supply (20).

EXAMPLE 25

Primer-Directed And Primer Independent Cleavage Occur At The Same Site When The Primer Extends To The 3' Side Of A Mismatched "Bubble" In The Downstream Duplex

As discussed above in Example 1, the presence of a primer upstream of a bifurcated duplex can influence the site of cleavage, and the existence of a gap between the 3' end of the primer and the base of the duplex can cause a shift of the cleavage site up the unpaired 5' arm of the structure (see also Lyamichev *et al., supra* and U.S. Patent No. 5,422,253). The resulting non-invasive shift of the cleavage site in response to a primer is demonstrated in Figs. 8, 9 and 10, in which the primer used left a 4-nucleotide gap (relative to the base of the duplex). In Figs. 8-10, all of the "primer-directed" cleavage reactions yielded a 21 nucleotide product, while the primer-independent cleavage reactions yielded a 25 nucleotide product. The site of cleavage obtained when the primer was extended to the base of the duplex, leaving no gap was examined. The results are shown in Fig. 53 (Fig. 53 is a reproduction of Fig. 2C in Lyamichev *et al.* These data were derived from the cleavage of the structure shown in Fig. 5, as described in Example 1. Unless otherwise specified, the cleavage reactions comprised 0.01 pmoles of heat-denatured, end-labeled hairpin DNA (with the unlabeled complementary strand also present), 1 pmole primer (complementary to the 3' arm shown in Fig. 5 and having the sequence: 5'-GAATTCGATTTAGGTGACAC TATAGAATACA [SEQ ID NO:53]) and 0.5 units of DNAP*Taq* (estimated to be 0.026 pmoles) in a total volume of 10 μl of 10 mM Tris-Cl, pH 8.5, and 1.5 mM $MgCl_2$ and 50 mM KCl. The primer was omitted from the reaction shown in the first lane of Fig. 53 and included in lane 2. These reactions were incubated at 55°C for 10 minutes. Reactions were initiated at the final reaction temperature by the addition of either the $MgCl_2$ or enzyme. Reactions were stopped at their incubation temperatures by the addition of 8 μl of 95% formamide with 20 mM EDTA and 0.05% marker dyes.

Fig. 53 is an autoradiogram that indicates the effects on the site of cleavage of a bifurcated duplex structure in the presence of a primer that extends to the base of the hairpin duplex. The size of the released cleavage product is shown to the left (*i.e.*, 25 nucleotides).

A dideoxynucleotide sequencing ladder of the cleavage substrate is shown on the right as a marker (lanes 3-6).

These data show that the presence of a primer that is adjacent to a downstream duplex (lane 2) produces cleavage at the same site as seen in reactions performed in the absence of the primer (lane 1). (*See* Figs. 8A and B, 9B and 10A for additional comparisons). When the 3' terminal nucleotides of the upstream oligonucleotide can base pair to the template strand but are not homologous to the displaced strand in the region immediately upstream of the cleavage site (*i.e.*, when the upstream oligonucleotide is opening up a "bubble" in the duplex), the site to which cleavage is apparently shifted is not wholly dependent on the presence of an upstream oligonucleotide.

As discussed above in the Background, and in Table 1, the requirement that two independent sequences be recognized in an assay provides a highly desirable level of specificity. In the invasive cleavage reactions of the present invention, the Invader™ and probe oligonucleotides must hybridize to the target nucleic acid with the correct orientation and spacing to enable the production of the correct cleavage product. When the distinctive pattern of cleavage is not dependent on the successful alignment of both oligonucleotides in the detection system these advantages of independent recognition are lost.

EXAMPLE 26

Invasive Cleavage And Primer-Directed Cleavage When There Is Only Partial Homology In The "X" Overlap Region

While not limiting the present invention to any particular mechanism, invasive cleavage occurs when the site of cleavage is shifted to a site within the duplex formed between the probe and the target nucleic acid in a manner that is dependent on the presence of an upstream oligonucleotide which shares a region of overlap with the downstream probe oligonucleotide. In some instances, the 5' region of the downstream oligonucleotide may not be completely complementary to the target nucleic acid. In these instances, cleavage of the probe may occur at an internal site within the probe even in the absence of an upstream oligonucleotide (in contrast to the base-by-base nibbling seen when a fully paired probe is used without an Invader™). Invasive cleavage is characterized by an apparent shifting of cleavage to a site within a downstream duplex that is dependent on the presence of the Invader™ oligonucleotide.

A comparison between invasive cleavage and primer-directed cleavage may be illustrated by comparing the expected cleavage sites of a set of probe oligonucleotides having decreasing degrees of complementarity to the target strand in the 5' region of the probe (*i.e.*, the region that overlaps with the Invader™). A simple test, similar to that performed on the hairpin substrate above (Ex. 25), can be performed to compare invasive cleavage with the non-invasive primer-directed cleavage described above. Such a set of test oligonucleotides is diagrammed in Fig. 54. The structures shown in Fig. 54 are grouped in pairs, labeled "a", "b", "c", and "d". Each pair has the same probe sequence annealed to the target strand (SEQ ID NO:54), but the top structure of each pair is drawn without an upstream oligonucleotide, while the bottom structure includes this oligonucleotide (SEQ ID NO:55). The sequences of the probes shown in Figs. 54a-54d are listed in SEQ ID NOS:32, 56, 57 and 58, respectively. Probable sites of cleavage are indicated by the black arrowheads. (It is noted that the precise site of cleavage on each of these structures may vary depending on the choice of cleavage agent and other experimental variables. These particular sites are provided for illustrative purposes only.)

To conduct this test, the site of cleavage of each probe is determined both in the presence and the absence of the upstream oligonucleotide, in reaction conditions such as those described in Example 18. The products of each pair of reactions are then be compared to determine whether the fragment released from the 5' end of the probe increases in size when the upstream oligonucleotide is included in the reaction.

The arrangement shown in Fig. 54a, in which the probe molecule is completely complementary to the target strand, is similar to that shown in Fig. 28. Treatment of the top structure with the 5' nuclease of a DNA polymerase would cause exonucleolytic nibbling of the probe (*i.e.*, in the absence of the upstream oligonucleotide). In contrast, inclusion of an Invader™ oligonucleotide would cause a distinctive cleavage shift similar, to those observed in Fig. 29.

The arrangements shown in Figs. 54b and 54c have some amount of unpaired sequence at the 5' terminus of the probe (3 and 5 bases, respectively). These small 5' arms are suitable cleavage substrate for the 5' nucleases and would be cleaved within 2 nucleotide's of the junction between the single stranded region and the duplex. In these arrangements, the 3' end of the upstream oligonucleotide shares identity with a portion of the 5' region of the probe which is complementary to the target sequence (that is the 3' end of the Invader™ has to compete for binding to the target with a portion of the 5' end of the

probe). Therefore, when the upstream oligonucleotide is included it is thought to mediate a shift in the site of cleavage into the downstream duplex (although the present invention is not limited to any particular mechanism of action), and this would, therefore, constitute invasive cleavage. If the extreme 5' nucleotides of the unpaired region of the probe were able to hybridize to the target strand, the cleavage site in the absence of the Invader™ might change but the addition of the Invader™ oligonucleotide would still shift the cleavage site to the proper position.

Finally, in the arrangement shown in Fig. 54d, the probe and upstream oligonucleotides share no significant regions of homology, and the presence of the upstream oligonucleotide would not compete for binding to the target with the probe. Cleavage of the structures shown in Fig. 54d would occur at the same site with or without the upstream oligonucleotide, and is thus would not constitute invasive cleavage.

By examining any upstream oligonucleotide/probe pair in this way, it can easily be determined whether the resulting cleavage is invasive or merely primer-directed. Such analysis is particularly useful when the probe is not fully complementary to the target nucleic acid, so that the expected result may not be obvious by simple inspection of the sequences.

EXAMPLE 27

Modified Cleavase® Enzymes

In order to develop nucleases having useful activities for the cleavage of nucleic acids the following modified nucleases were produced.

a) Cleavase® BN/thrombin Nuclease

i) Cloning and Expression of Cleavase® BN/thrombin Nuclease

Site directed mutagenesis was used to introduce a protein sequence recognized by the protease thrombin into the region of the Cleavase® BN nuclease which is thought to form the helical arch of the protein through which the single-stranded DNA that is cleaved must presumably pass. Mutagenesis was carried out using the Transformer™ mutagenesis kit (Clonetech, Palo Alto, CA) according to manufacturer's protocol using the mutagenic oligonucleotide 5'-GGGAAAGTCCTCGCAGCCGCGCG GGACGAGCGTGGGGGCCCG (SEQ ID NO:59). After mutagenesis, the DNA was sequenced to verify the insertion of the thrombin cleavage site. The DNA sequence encoding the Cleavase® BN/thrombin nuclease is

provided in SEQ ID NO:60; the amino acid sequence of Cleavase® BN/thrombin nuclease is provided in SEQ ID NO:61.

A large scale preparation of the thrombin mutant (*i.e.*, Cleavase® BN/thrombin) was done using *E. coli* cells overexpressing the Cleavase® BN/thrombin nuclease as described in Example 28.

ii) Thrombin Cleavage of Cleavase® BN/thrombin

Six point four (6.4) mg of the purified Cleavase® BN/thrombin nuclease was digested with 0.4 U of thrombin (Novagen) for 4 hours at 23°C or 37°C. Complete digestion was verified by electrophoresis on a 15% SDS polyacrylamide gel followed by staining with Coomassie Brilliant Blue R. Wild-type Cleavase® BN nuclease was also digested with thrombin as a control. The resulting gel is shown in Fig. 61.

In Fig. 61, lane 1 contains molecular weight markers (Low-Range Protein Molecular Weight Markers; Promega), lane 2 contains undigested Cleavase® BN/throbin nuclease, lanes 3 and 4 contain Cleavase® BN/thrombin nuclease digested with thrombin at 23°C for 2 and 4 hours, respectively, and lanes 5 and 6 contain Cleavase® BN/thrombin nuclease digested with thrombin at 37°C for 2 and 4 hours, respectively. These results show that the Cleavase® BN/thrombin nuclease has an apparent molecular weight of 36.5 kilodaltons and demonstrate that Cleavase® BN/thrombin nuclease is efficiently cleaved by thrombin. In addition, the thrombin cleavage products have approximate molecular weights of 27 kilodaltons and 9 kilodaltons, the size expected based upon the position of the inserted thrombin site in the Cleavase® BN/thrombin nuclease.

To determine the level of hairpin cleavage activity in digested and undigested Cleavase® BN/thrombin nuclease, dilutions were made and used to cleave a test hairpin containing a 5' fluoroscein label. Varying amounts of digested and undigested Cleavase® BN/thrombin nuclease were incubated with 5 μM oligonucleotide S-60 hairpin (SEQ ID NO:29; see Fig. 26) in 10 mM MOPS (pH 7.5), 0.05% Tween-20, 0.05% NP-40, and 1 mM $MnCl_2$ for 5 minutes at 60°C. The digested mixture was electrophoresed on a 20% acrylamide gel and visualized on a Hitachi FMBIO 100 fluoroimager. The resulting image is shown in Fig. 62.

In Fig. 62, lane 1 contains the no enzyme control, lane 2 contains reaction products produced using 0.01 ng of Cleavase® BN nuclease, lanes 3, 4, and 5 contain reaction products produced using 0.01 ng, 0.04 ng, and 4 ng of undigested Cleavase® BN/thrombin

nuclease, respectively, and lanes 6, 7, and 8 contain reaction products produced using 0.01 ng, 0.04 ng, and 4 ng of thrombin-digested Cleavase® BN/thrombin nuclease, respectively. The results shown in Fig. 62 demonstrated that the insertion of the thrombin cleavage site reduced cleavage activity about 200-fold (relative to the activity of Cleavase® BN nuclease), but that digestion with thrombin did not reduce the activity significantly.

M13 single-stranded DNA was used as a substrate for cleavage by Cleavase® BN nuclease and digested and undigested Cleavase® BN/thrombin nuclease. Seventy nanograms of single-stranded M13 DNA (NEB) was incubated in 10 mM MOPS, pH 7.5, 0.05% Tween-20, 0.05% NP-40, 1 mM $MgCl_2$ or 1 mM $MnCl_2$, with 8 ng of Cleavase® BN nuclease, undigested Cleavase® BN/thrombin nuclease, or digested Cleavase® BN/thrombin nuclease for 10 minutes at 50°C. Reaction mixtures were electrophoresed on a 0.8% agarose gel and then stained with a solution containing 0.5 μg/ml ethidium bromide (EtBr) to visualize DNA bands. A negative image of the EtBr-stained gel is shown in Fig. 63.

In Fig. 63, lane 1 contains the no enzyme control, lane 2 contains reaction products produced using Cleavase® BN nuclease and 1 mM $MgCl_2$, lane 3 contains reaction products produced using Cleavase® BN nuclease and 1 mM $MnCl_2$, lane 4 contains reaction products produced using undigested Cleavase® BN/thrombin nuclease and 1 mM $MgCl_2$, lane 5 contains reaction products produced using undigested Cleavase® BN/thrombin nuclease and 1 mM $MnCl_2$, lane 6 contains reaction products produced using thrombin-digested Cleavase® BN/thrombin nuclease and 1 mM $MgCl_2$, and lane 7 contains reaction products produced using thrombin-digested Cleavase® BN/thrombin nuclease and 1 mM $MnCl_2$. The results shown in Fig. 63 demonstrated that the Cleavase® BN/thrombin nuclease had an enhanced ability to cleave circular DNA (and thus a reduced requirement for the presence of a free 5' end) as compared to the Cleavase® BN nuclease.

It can be seen from these data that the helical arch of these proteins can be opened without destroying the enzyme or its ability to specifically recognize cleavage structures. The Cleavase® BN/thrombin mutant has an increased ability to cleave without reference to a 5' end, as discussed above. The ability to cleave such structures will allow the cleavage of long molecules, such as genomic DNA that, while often not circular, may present many desirable cleavage sites that are at a far removed from any available 5' end. Cleavage structures may be made at such sites either by folding of the strands (*i.e.*, CFLP® cleavage) or by the introduction of structure-forming oligonucleotides (U.S. Patent No. 5,422,253). 5' ends of nucleic acids can also be made unavailable because of binding of a substance too large to

thread through the helical arch. Such binding moieties may include proteins such as streptavidin or antibodies, or solid supports such as beads or the walls of a reaction vessel. A cleavage enzyme with an opening in the loop of the helical arch will be able to cleave DNAs that are configured in this way, extending the number of ways in which reactions using such enzymes can be formatted.

b) Cleavase® DN Nuclease

i) Construction and Expression of Cleavase® DN Nuclease

A polymerization deficient mutant of *Taq* DNA polymerase, termed Cleavase® DN nuclease, was constructed. Cleavase® DN nuclease contains an asparagine residue in place of the wild-type aspartic acid residue at position 785 (D785N).

DNA encoding the Cleavase® DN nuclease was constructed from the gene encoding for Cleavase® A/G (mut*Taq*, Ex. 2) in two rounds of site-directed mutagenesis. First, the G at position 1397 and the G at position 2264 of the Cleavase® A/G gene (SEQ ID NO:21) were changed to A at each position to recreate a wild-type DNAPTaq gene. As a second round of mutagenesis, the wild type DNAPTaq gene was converted to the Cleavase® DN gene by changing the G at position 2356 to A. These manipulations were performed as follows.

DNA encoding the Cleavase® A/G nuclease was recloned from pTTQ18 plasmid (Ex. 2) into the pTrc99A plasmid (Pharmacia) in a two step procedure.
First, the pTrc99A vector was modified by removing the G at position 270 of the pTrc99A map, creating the pTrc99G cloning vector. To this end, pTrc99A plasmid DNA was cut with *Nco*I and the recessive 3' ends were filled-in using the Klenow fragment of *E.coli* polymerase I in the presence of all four dNTPs at 37°C for 15 min. After inactivation of the Klenow fragment by incubation at 65°C for 10 min, the plasmid DNA was cut with *Eco*RI, the ends were again filled-in using the Klenow fragment in the presence of all four dNTPs at 37°C for 15 min. The Klenow fragment was then inactivated by incubation at 65°C for 10 min. The plasmid DNA was ethanol precipitated, recircularized by ligation, and used to transform *E.coli* JM109 cells (Promega). Plasmid DNA was isolated from single colonies and deletion of the G at position 270 of the pTrc99A map was confirmed by DNA sequencing.

As a second step, DNA encoding the Cleavase® A/G nuclease was removed from the pTTQ18 plasmid using *Eco*RI and *Sal*I and the DNA fragment carrying the Cleavase® A/G nuclease gene was separated on a 1% agarose gel and isolated with Geneclean II Kit (Bio

101, Vista, CA). The purified fragment was ligated into the pTrc99G vector which had been cut with *Eco*RI and *Sal*I. The ligation mixture was used to transform competent *E.coli* JM109 cells (Promega). Plasmid DNA was isolated from single colonies and insertion of the Cleavase® A/G nuclease gene was confirmed by restriction analysis using *Eco*RI and *Sal*I.

Plasmid DNA pTrcAG carrying the Cleavase® A/G nuclease gene cloned into the pTrc99A vector was purified from 200 ml of JM109 overnight culture using QIAGEN Plasmid Maxi kit (QIAGEN, Chatsworth,CA) according to manufacturer's protocol. pTrcAG plasmid DNA was mutagenized using two mutagenic primers, E465 (SEQ ID NO:62) (Integrated DNA Technologies, Iowa) and R754Q (SEQ ID NO:63) (Integrated DNA Technologies), and the selection primer Trans Oligo AlwNI/SpeI (Clontech, Palo Alto, CA, catalog #6488-1) according to Transformer™ Site-Directed Mutagenesis Kit protocol (Clontech) to produce a restored wild-type DNAPTaq gene (pTrcWT).

pTrcWT plasmid DNA carrying the wild-type DNAPTaq gene cloned into the pTrc99A vector was purified from 200 ml of JM109 overnight culture using QIAGEN Plasmid Maxi kit (QIAGEN, Chatsworth, CA) according to manufacturer's protocol. pTrcWT was then mutagenized using the mutagenic primer D785N (SEQ ID NO:64) (Integrated DNA Technologies) and the selection primer Switch Oligo SpeI/AlwNI (Clontech, catalog #6373-1) according to Transformer™ Site-Directed Mutagenesis Kit protocol (Clontech) to create a plasmid containing DNA encoding the Cleavase® DN nuclease. The DNA sequence encoding the Cleavase® DN nuclease is provided in SEQ ID NO:65; the amino acid sequence of Cleavase® DN nuclease is provided in SEQ ID NO:66.

A large scale preparation of the Cleavase® DN nuclease was done using *E. coli* cells overexpressing the Cleavase® DN nuclease as described in Example 28.

c) Cleavase® DA Nuclease and Cleavase® DV Nuclease

Two polymerization deficient mutants of *Taq* DNA polymerase, termed Cleavase® DA nuclease and Cleavase® DV nuclease, were constructed. The Cleavase® DA nuclease contains a alanine residue in place of the wild-type aspartic acid residue at position 610 (D785A). The Cleavase® DV nuclease contains a valine residue in place of the wild-type aspartic acid residue at position 610 (D610V).

i) Construction and Expression of the Cleavase® DA and Cleavase® DV Nucleases

To construct vectors encoding the Cleavase® DA and DV nucleases, the Cleavase® A/G nuclease gene contained within pTrcAG was mutagenized with two mutagenic primers, R754Q (SEQ ID NO:63) and D610AV (SEQ ID NO:67) and the selection primer Trans Oligo AlwNI/SpeI (Clontech, catalog #6488-1) according to the Transformer™ Site-Directed Mutagenesis Kit protocol (Clontech,) to create a plasmid containing DNA encoding the Cleavase® DA nuclease or Cleavase® DV nuclease. The D610AV oligonucleotide was synthesized to have a purine, A or G, at position 10 from the 5' end of the oligonucleotide. Following mutagenesis, plasmid DNA was isolated from single colonies and the type of mutation present, DA or DV, was determined by DNA sequencing. The DNA sequence encoding the Cleavase® DA nuclease is provided in SEQ ID NO:68; the amino acid sequence of Cleavase® DA nuclease is provided in SEQ ID NO:69. The DNA sequence encoding the Cleavase® DV nuclease is provided in SEQ ID NO:70; the amino acid sequence of Cleavase® DV nuclease is provided in SEQ ID NO:71.

Large scale preparations of the Cleavase® DA and Cleavase® DV nucleases was done using *E. coli* cells overexpressing the Cleavase® DA nuclease or the Cleavase® DV nuclease as described in Example 28.

EXAMPLE 28

Cloning And Expression of Thermostable FEN-1 Endonucleases

Sequences encoding thermostable FEN-1 proteins derived from three Archaebacterial species were cloned and overexpressed in *E. coli.* This Example involved a) cloning and expression of a FEN-1 endonuclease from *Methanococcus jannaschii*; b) cloning and expression of a FEN-1 endonuclease from *Pyrococcus furiosus*; c) cloning and expression of a FEN-1 endonuclease from *Pyrococcus woesei*; d) cloning and expression of a FEN-1 endonuclease from *Archaeoglobus fulgidus*; e) large scale preparation of recombinant thermostable FEN-1 proteins; and f) activity assays using FEN-1 endonucleases.

a) Cloning and Expression Of A FEN-1 Endonuclease From *Methanococcus jannaschii*

DNA encoding the FEN-1 endonuclease from *Methanococcus jannaschii* (*M. jannaschii*) was isolated from *M. jannaschii* cells and inserted into a plasmid under the transcriptional control of an inducible promoter as follows. Genomic DNA was prepared from 1 vial of live *M. jannaschii* bacteria (DSMZ, Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany # 2661) with the DNA XTRAX kit (Gull Laboratories, Salt Lake City, UT) according to the manufacturer's protocol. The final DNA pellet was resuspended in 100 μl of TE (10 mM Tris HCl, pH 8.0, 1 mM EDTA). One microliter of the DNA solution was employed in a PCR using the Advantage™ cDNA PCR kit (Clonetech); the PCR was conducted according to manufacturer's recommendations. The 5'-end primer (SEQ ID NO:72) is complementary to the 5' end of the *Mja* FEN-1 open reading frame with a one base substitution to create an *Nco*I restriction site (a fragment of the *M. jannaschii* genome which contains the gene encoding *M. jannaschii* (*Mja*) FEN-1 is available from GenBank as accession # U67585). The 3'-end primer (SEQ ID NO:73) is complementary to a sequence about 15 base pairs downstream from the 3' end of the *Mja* FEN-1 open reading frame with 2 base substitutions to create a *Sal*I restriction enzyme site. The sequences of the 5'-end and 3'-end primers are: 5'-GGGATACCA TGGGAGTGCAGTTTGG-3' (SEQ ID NO:72) and 5'-GGTAAATTTTTCTCGTCGA CATCCCAC-3' (SEQ ID NO:73), respectively. The PCR reaction resulted in the amplification (*i.e.*, production) of a single major band about 1 kilobase in length. The open reading frame (ORF) encoding the *Mja* FEN-1 endonuclease is provided in SEQ ID NO:74; the amino acid sequence encoded by this ORF is provided in SEQ ID NO:75.

Following the PCR amplification, the entire reaction was electrophoresed on a 1.0% agarose gel and the major band was excised from the gel and purified using the Geneclean II kit (Bio101, Vista, CA) according to manufacturer's instructions. Approximately 1 μg of the gel-purified *Mja* FEN-1 PCR product was digested with *Nco*I and *Sal*I. After digestion, the DNA was purified using the Geneclean II kit according to manufacturer's instructions. One microgram of the pTrc99a vector (Pharmacia) was digested with *Nco*I and *Sal*I in preparation for ligation with the digested PCR product. One hundred nanograms of digested pTrc99a vector and 250 ng of digested *Mja* FEN-1 PCR product were combined and ligated to create pTrc99-MJFEN1. pTrc99-MJFEN1 was used to transform competent *E. coli* JM109 cells (Promega) using standard techniques.

b) Cloning and Expression Of A FEN-1 Endonuclease From *Pyrococcus furiosus*

DNA encoding the *Pyrococcus furiosus* (*P. furiosus*) FEN-1 endonuclease was obtained by PCR amplification using a plasmid containing DNA encoding the *P. furiosus* (*Pfu*) FEN-1 endonuclease (obtained from Dr. Frank Robb, Center of Marine Biotechnology, Baltimore, MD). DNA sequences encoding a portion of the *Pfu* FEN-1 endonuclease can be obtained from GenBank as accession Nos. AA113505 and W36094. The amplified *Pfu* FEN-1 gene was inserted into the pTrc99a expression vector (Pharmacia) to place the *Pfu* FEN-1 gene under the transcriptional control of the inducible *trc* promoter. The PCR amplification was conducted as follows. One hundred microliter reactions contained 50 mM Tris HCl, pH 9.0, 20 mM $(NH_4)_2SO_4$, 2 mM $MgCl_2$, 50 μM dNTPs, 50 pmole each primer, 1 U *Tfl* polymerase (Epicentre Technologies, Madison, WI) and 1 ng of FEN-1 gene-containing plasmid DNA. The 5'-end primer (SEQ ID NO:76) is complementary to the 5' end of the *Pfu* FEN-1 open reading frame but with two substitutions to create an *Nco*I site and the 3'-end primer (SEQ ID NO:77) is complementary to a region located about 30 base pairs downstream of the FEN-1 open reading frame with two substitutions to create a *Pst*I site. The sequences of the 5'-end and 3'-end primers are: 5'-GAGGTGATACCATG GGTGTCC-3' (SEQ ID NO:76) and 5'-GAAACTCTGCAGCGCGTCAG-3' (SEQ ID NO:77), respectively. The PCR reaction resulted in the amplification of a single major band about 1 kilobase in length. The open reading frame (ORF) encoding the *Pfu* FEN-1 endonuclease is provided in SEQ ID NO:78; the amino acid sequence encoded by this ORF is provided in SEQ ID NO:79.

Following the PCR amplification, the entire reaction was electrophoresed on a 1.0% agarose gel and the major band was excised from the gel and purified using the Geneclean II kit (Bio101) according to manufacturer's instructions. Approximately 1 μg of gel purified *Pfu* FEN-1 PCR product was digested with *Nco*I and *Pst*I. After digestion, the DNA was purified using the Geneclean II kit according to manufacturer's instructions. One microgram of the pTrc99a vector was digested with *Nco*I and *Pst*I prior to ligation with the digested PCR product. One hundred nanograms of digested pTrc99a and 250 ng of digested *Pfu* FEN-1 PCR product were combined and ligated to create pTrc99-PFFEN1. pTrc99-PFFEN1 was used to transform competent *E. coli* JM109 cells (Promega) using standard techniques.

c) **Cloning and Expression Of A FEN-1 Endonuclease From *Pyrococcus woesei***

For the cloning of DNA encoding the *Pyrococcus woesei* (*Pwo*) FEN-1 endonuclease, DNA was prepared from lyophilized *P. woesei* bacteria (DSMZ # 3773) as described (Zwickl *et al.*, J. Bact., 172:4329 [1990]) with several changes. Briefly, one vial of *P. woesei* bacteria was rehydrated and resuspended in 0.5 ml of LB (Luria broth). The cells were centrifuged at 14,000 x *g* for 1 min and the cell pellet was resuspended in 0.45 ml of TE. Fifty microliters of 10% SDS was added and the mixture was incubated at RT for 5 min. The cell lysate was then extracted three time with 1:1 phenol:chloroform and three times with chloroform. Five hundred microliters of isopropanol was added to the extracted lysate and the DNA was pelleted by centrifugation at 14,000 x *g* for 10 min. The DNA pellet was washed in 0.5 ml of 70% ethanol and the DNA was pelleted again by centrifugation at 14,000 x *g* for 5 min. The DNA pellet was dried and resuspended in 100 µl of TE and used for PCR reactions without further purification.

To generate a *P. woesei* FEN-1 gene fragment for cloning into an expression vector, low stringency PCR was attempted with primers complementary to the ends of the *P. furiosus* FEN-1 gene open reading frame. The sequences of the 5'-end and 3'-end primers are 5'-GATACCATGGGTGTCCCAATTGGTG-3' (SEQ ID NO:80) and 5'-TCGACGTCGACTTATCTCTTGAACCAACTTTCAAGGG-3' (SEQ ID NO:81), respectively. The high level of sequence similarity of protein homologs (*i.e.*, proteins other than FEN-1 proteins) from *P. furiosus* and *P. woesei* suggested that there was a high probability that the *P. woesei* FEN-1 gene could be amplified using primers containing sequences complementary to the *P. furiosus* FEN-1 gene. However, this approach was unsuccessful under several different PCR conditions.

The DNA sequence of FEN-1 genes from *P. furiosus* and *M. jannaschii* were aligned and blocks of sequence identity between the two genes were identified. These blocks were used to design internal primers (*i.e.*, complementary to sequences located internal to the 5' and 3' ends of the ORF) for the FEN-1 gene that are complementary to the *P. furiosus* FEN-1 gene in those conserved regions. The sequences of the 5'- and 3'-internal primers are 5'-AGCGAGGGAGAGGCCCAAGC-3' (SEQ ID NO:82) and 5'-GCCTATGCCCTTTATTCCTCC-3' (SEQ ID NO:83), respectively. A PCR employing these internal primers was conducted using the Advantage™ PCR kit and resulted in production of a major band of ~300 bp.

Since the PCR with the internal primers was successful, reactions were attempted which contained mixtures of the internal (SEQ ID NOS:82 and 83) and external (SEQ ID NOS:80 and 81) primers. A reaction containing the 5'-end external primer (SEQ ID NO:80) and 3'-end internal primer (SEQ ID NO:83) resulted in the production of a 600 bp band and a reaction containing the 5'-end internal primer (SEQ ID NO:82) and 3'-end external primer (SEQ ID NO:81) resulted in the production of a 750 bp band. These overlapping DNA fragments were gel-purified and combined with the external primers (SEQ ID NOS:80 and 81) in a PCR reaction. This reaction generated a 1 kb DNA fragment containing the entire *Pwo* FEN-1 gene open reading frame. The resulting PCR product was gel-purified, digested, and ligated exactly as described above for the *Mja* FEN-1 gene PCR product. The resulting plasmid was termed pTrc99-PWFEN1. pTrc99-PWFEN1 was used to transform competent *E. coli* JM109 cells (Promega) using standard techniques.

d) Cloning and Expression Of A FEN-1 Endonuclease From *Archaeoglobus fulgidus*

The preliminary *Archaeoglobus fulgidus* (Afu) chromosome sequence of 2.2 million bases was downloaded from the TIGR (The Institute for Genomic Research) world wide web site, and imported into a software program (MacDNAsis), used to analyze and manipulate DNA and protein sequences. The unannotated sequence was translated into all 6 of the possible reading frames, each comprising approximately 726,000 amino acids. Each frame was searched individually for the presence of the amino acid sequence "VFDG" (valine, phenylalanine, aspartic acid, glycine), a sequence which is conserved in the FEN-1 family. The amino acid sequence was found in an open reading frame that contained other amino acid sequences conserved in the FEN-1 genes and which was approximately the same size as the other FEN-1 genes. The ORF DNA sequence is shown in SEQ ID NO:164, while the ORF protein sequence is shown in SEQ ID NO:165. Based on the position of this amino acid sequence within the reading frame, the DNA sequence encoding a putative FEN-1 gene was identified.

The sequence information was used to design oligonucleotide primers which were used for PCR amplification of the FEN-1-like sequence from *A. fulgidus* genomic DNA. Genomic DNA was prepared from *A. fulgidus* as described in Ex. 29a for *M. janaschii*, except that one vial (approximately 5 ml of culture) of live *A. fulgidus* bacteria from DSMZ (DSMZ #4304) was used. One microliter of the genomic DNA was used for PCR reaction as described in

Ex. 29a. The 5' end primer is complementary to the 5' end of the Afu FEN-1 gene except it has a 1 base pair substitution to create an *Nco* I site. The 3' end primer is complentary to the 3' end of the Afu FEN-1 gene downstream from the FEN-1 ORF except it contains a 2 base substitution to create a Sal I site. The sequences of the 5' and 3' end primers are 5'-CCGTCAACATTTACCATGGGTGCGGA-3' (SEQ ID NO:166) and 5'-CCGCCACCTCGTAGTCGACATCCTTTTCGTG (SEQ ID NO:167), respectively.

Cloning of the resulting fragment was as described for the PfuFEN1 gene, above, to create the plasmid pTrc99-AFFEN1. The pTrcAfuHis plasmid was constructed by modifying pTrc99-AFFEN1, by adding a histidine tail to facilitate purification. To add this histidine tail, standard PCR primer-directed mutagenesis methods were used to insert the coding sequence for six histidine residues between the last amino acid codon of the pTrc99-AFFEN1 coding region and the stop codon. The resulting plasmid was termed pTrcAfuHis. The protein was then expressed as described in Example 28(e), and purified by binding to a Ni++ affinity column, as described in Example 8.

e) Large Scale Preparation of Recombinant Thermostable FEN-1 Proteins

The *Mja*, *Pwo* and *Pfu* FEN-1 proteins were purified by the following technique which is derived from a *Taq* DNA polymerase preparation protocol (Engelke *et al.*, Anal. Biochem., 191:396 [1990]) as follows. *E. coli* cells (strain JM109) containing either pTrc99-PFFEN1, pTrc99-PWFEN1, or pTrc99-MJFEN1 were inoculated into 3 ml of LB (Luria Broth) containing 100 μg/ml ampicillin and grown for 16 hrs at 37°C. The entire overnight culture was inoculated into 200 ml or 350 ml of LB containing 100 μg/ml ampicillin and grown at 37°C with vigorous shaking to an A_{600} of 0.8. IPTG (1 M stock solution) was added to a final concentration of 1 mM and growth was continued for 16 hrs at 37°C.

The induced cells were pelleted and the cell pellet was weighed. An equal volume of 2X DG buffer (100 mM Tris-HCl, pH 7.6, 0.1 mM EDTA) was added and the pellet was resuspended by agitation. Fifty mg/ml lysozyme (Sigma, St. Louis, MO) was added to 1 mg/ml final concentration and the cells were incubated at room temperature for 15 min. Deoxycholic acid (10% solution) was added dropwise to a final concentration of 0.2 % while vortexing. One volume of H_2O and 1 volume of 2X DG buffer was added and the resulting mixture was sonicated for 2 minutes on ice to reduce the viscosity of the mixture. After

sonication, 3 M $(NH_4)_2SO_4$ was added to a final concentration of 0.2 M and the lysate was centrifuged at 14000 x *g* for 20 min at 4°C. The supernatant was removed and incubated at 70°C for 60 min at which time 10% polyethylimine (PEI) was added to 0.25%. After incubation on ice for 30 min., the mixture was centrifuged at 14,000 x *g* for 20 min at 4°C. At this point, the supernatant was removed and the FEN-1 proteins was precipitated by the addition of $(NH_4)_2SO_4$ as follows.

For the *Pwo* and the *Pfu* FEN-1 preparations, the FEN-1 protein was precipitated by the addition of 2 volumes of 3 M $(NH_4)_2SO_4$. The mixture was incubated overnight at room temperature for 16 hrs and the protein was centrifuged at 14,000 x *g* for 20 min at 4°C. The protein pellet was resuspended in 0.5 ml of Q buffer (50 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, 0.1% Tween 20). For the *Mja* FEN-1 preparation, solid $(NH_4)_2SO_4$ was added to a final concentration of 3 M (~75% saturated), the mixture was incubated on ice for 30 min, and the protein was spun down and resuspended as described above.

The resuspended protein preparations were quantitated by determination of the A_{279} and aliquots containing 2-4 μg of total protein were electrophoresed on a 10 % SDS polyacrylamide gel (29:1 acrylamide: bis-acrylamide) in standard Laemmli buffer [Laemmli, Nature 277:680 [1970]) and stained with Coomassie Brilliant Blue R; the results are shown in Fig. 64.

In Fig. 64, lane 1 contains molecular weight markers (Mid-Range Protein Molecular Weight Markers; Promega); the size of the marker proteins is indicated to the left of the gel. Lane 2 contains purified Cleavase® BN nuclease; lanes 3-5 contain extracts prepared from *E. coli* expressing the *Pfu*, *Pwo* and *Mja* FEN-1 nucleases, respectively. The calculated (*i.e.*, using a translation of the DNA sequence encoding the nuclease) molecular weight of the *Pfu* FEN-1 nuclease is 38,714 daltons and the calculated molecular weight for the *Mja* FEN-1 nuclease is 37,503 Daltons. The *Pwo* and *Pfu* FEN-1 proteins co-migrated on the SDS-PAGE gel and therefore, the molecular weight of the *Pwo* FEN-1 nuclease was estimated to be 38.7 kDa.

f) Activity Assays Using FEN-1 Endonucleases

i) Mixed Hairpin Assay

The Cleavase® BN nuclease has an approximately 60-fold greater affinity for a 12 base pair stem-loop structure than an 8 base pair stem-loop DNA structure. As a test for activity differences between the Cleavase® BN nuclease and the FEN-1 nucleases, a mixture

of oligonucleotides having either a 8 or a 12 bp stem-loop (see Fig. 60 which depicts the S-33 and 11-8-0 oligonucleotides) was incubated with an extract prepared from *E. coli* cells overexpressing the *Mja* FEN-1 nuclease (prepared as described above). Reactions contained 0.05 μM of oligonucleotides S-33 (SEQ ID NO:84) and 11-8-0 (SEQ ID NO:85) (both oligonucleotides contained 5'-fluorescein labels), 10 mM MOPS, pH 7.5, 0.05% Tween-20, 0.05% NP-40, 1 mM $MnCl_2$. Reactions were heated to 90°C for 10 seconds, cooled to 55°C, then 1 μl of crude extract (*Mja* FEN-1) or purified enzyme (Cleavase® BN nuclease) was added and the mixtures were incubated at 55°C for 10 minutes; a no enzyme control was also run. The reactions were stopped by the addition of formamide/EDTA, the samples were electrophoresed on a denaturing 20% acrylamide gel and visualized on a Hitachi FMBIO 100 fluoroimager. The resulting image is shown in Fig. 65.

In Fig. 65, lane 1 contains the reaction products generated by the Cleavase® BN nuclease, lane 2 contains the reaction products from the no enzyme control reaction and lane 3 contains the reaction products generated by the *Mja* FEN-1 nuclease. The data shown in Fig. 76 demonstrates that the Cleavase® BN nuclease strongly prefers the S33 structure (12 bp stem-loop) while the *Mja* FEN-1 nuclease cleaves structures having either an 8 or a 12 bp stem-loop with approximately the same efficiency. This shows that the *Mja* FEN-1 nuclease has a different substrate specificity than the Cleavase® BN nuclease, a useful feature for Invader™ assays or CFLP® analysis as discussed in the Description of the Invention.

EXAMPLE 29

Terminal Deoxynucleotidyl Transferase Selectively Extends The Products Of Invader™-Directed Cleavage

The majority of thermal degradation products of DNA probes will have a phosphate at the 3'-end. To investigate if the template-independent DNA polymerase, terminal deoxynucleotide transferase (TdT) can tail or polymerize the aforementioned 3'-end phosphates (*i.e.*, add nucleotide triphosphates to the 3' end) the following experiment was performed.

To create a sample containing a large percentage of thermal degradation products, the 5' fluorescein-labelled oligonucleotide 34-078-01 (SEQ ID NO:86) (200 pmole) was incubated in 100 μl 10 mM $NaCO_3$ (pH 10.6), 50 mM NaCl at 95°C for 13 hours. To prevent evaporation, the reaction mixture was overlaid with 60 μl ChillOut™14 liquid wax. The reaction mixture was then divided into two equal aliquots (A and B). Aliquot A was

mixed with one-tenth volume 3M NaOAc followed by three volumes ethanol and stored at -20°C. Aliquot B was dephosphorylated by the addition of 0.5 µl of 1M $MgCl_2$ and 1 µl of 1unit/µl Calf Intestine Alkaline Phosphatase (CIAP) (Promega), with incubation at 37°C for 30 minutes. An equal volume of phenol:chloroform: isomayl alcohol (24:24:1) was added to the sample followed by vortexing for one minute and then centrifugation 5 minutes at maximum speed in a microcentrifuge to separate the phases. The aqueous phase was removed to a new tube to which one-tenth volume 3M NaOAc, and three volumes ethanol was added followed by storage at -20°C for 30 minutes. Both aliquots (A and B) were then centrifuged for 10 minutes at maximum speed in a microcentrifuge to pellet the DNA. The pellets were then washed two times each with 80% ethanol and then desiccated to dryness. The dried pellets were then dissolved in 70 µl ddH_2O each.

The TdT reactions were conducted as follows. Six mixes were assembled, all mixes contained 10 mM TrisOAc (pH 7.5), 10 mM MgOAc, 50 mM KCl, and 2 mM dATP. Mixes 1 and 2 contained one pmole of untreated 34-078-01 (SEQ ID NO:86), mixes 3 and 4 contained 2 µl of aliquot A (above), mixes 5 and 6 contained 2 µl of aliquot B (above). To each 9 µl of mixes 1,3 and 5, 1 µl ddH_2O was added, to each 9 µl of mixes 2, 4, and 6, 1 µl of 20 units/µl TdT (Promega) was added. The mixes were incubated at 37°C for 1 hour and then the reaction was terminated by the addition of 5 µl 95% formamide with 10 mM EDTA and 0.05% marker dyes. Five microliters of each mixture was resolved by electrophoresis through a 20% denaturing acrylamide gel (19:1 cross-linked) with 7 M urea, in a buffer containing 45 mM Tris-Borate (pH 8.3), 1.4 mM EDTA, and imaged using with the FMBIO Image Analyzer with a 505 nm filter. The resulting imager scan is shown in Fig. 66.

In Fig. 66, lanes 1, 3 and 5 contain untreated 34-078-01 (SEQ ID NO:86), heat-degraded 34-078-01, and heat-degraded, dephosphorylated, 34-078-01, respectively incubated in the absence of TdT. Lanes 2, 4 and 6 contain, untreated 34-078-01, heat-degraded 34-078-01, and heat-degraded, dephosphorylated, 34-078-01, respectively incubated in the presence of TdT.

As shown in Fig. 66, lane 4, TdT was unable to extend thermal degradation products which contain a 3'-end phosphate group, and selectively extends molecules which have a 3'-end hydroxyl group.

EXAMPLE 30

Specific TdT Tailing Of The Products Of Invader™-Directed Cleavage With Subsequent Capture And Detection On Nitrocellulose Supports

When TdT is used to extend the specific products of cleavage, one means of detecting the tailed products is to selectively capture the extension products on a solid support before visualization. This Example demonstrates that the cleavage products can be selectively tailed by the use of TdT and deoxynucleotide triphosphates, and that the tailed products can be visualized by capture using a complementary oligonucleotide bound to a nitrocellulose support.

To extend the cleavage product produced in an Invader™-directed cleavage reaction, the following experiment was performed. Three reaction mixtures were assembled, each in a buffer of 10 mM MES (pH 6.5), 0.5%Tween-20, 0.5% NP-40. The first mixture contained 5 fmols of target DNA-M13mp18, 10 pmols of probe oligo 32-161-2 (SEQ ID NO:87; this probe oligonucleotide contains 3' ddC and a Cy3 amidite group near the 3' end), and 5 pmols of Invader™ oligonucleotide 32 161-1 (SEQ ID NO:88; this oligo contains a 3' ddC). The second mixture contained the probe and Invader™ oligonucleotides without target DNA. The third mixture was the same as the first mixture, and contained the same probe sequence, but with a 5' fluorescein label (oligo 32-161-4 [SEQ ID NO:89; this oligo contains a 3' ddC, 5' fluorescein label, and a Cy3 dye group near the 3' end]), so that the Invader™-directed cleavage products could be detected before and after cleavage by fluorescence imaging. The probe only control sample contained 10 pmols of oligo 32-161-2 (SEQ ID NO:87). Each 3 μl of enzyme mix contained 5 ng of Cleavase® DN nuclease in 7.5 mM $MgCl_2$. The TdT mixture (per each 4 μl) contained: 10U of TdT (Promega), 1 mM $CoCl_2$, 50 mM KCl, and 100 μM of dTTP. The Invader™ cleavage reaction mixtures described above were assembled in thin wall tubes, and the reactions were initiated by the addition of 3 μl of Cleavase® DN enzyme mix. The reactions were incubated at 65°C for 20 min. After cooling to 37°C, 4 μl of the TdT mix was added and the samples were incubated for 4 min at 37°C, Biotin-16-dUTP was then added to 100 μM and the samples were incubated for 50 min at 37°C. The reactions were terminated by the addition of 1 μl of 0.5 M EDTA.

To test the efficiency of tailing the products were run on an acrylamide gel. Four microliters of each reaction mixture was mixed with 2.6 μl of 95% formamide, 10 mM EDTA and 0.05% methyl violet and heated to 90°C for 1 min, and 3 μl were loaded on a 20% denaturing acrylamide gel (19:1 cross-linked) with 7 M urea, in buffer containing 45

mM Tris-Borate (pH 8.3), 1.4 mM EDTA. A marker (ΦX174-*Hin*fI [fluorescein labeled]) also was loaded. After electrophoresis, the gel was analyzed using a FMBIO-100 Image Analyzer (Hitachi) equipped with a 505 nm filter. The resulting scan is shown in Fig. 67.

In Fig. 67, lane 1 contained the probe 32-161-2 only, without any treatment. Lanes 2 and 3 contained the products of reactions run without target DNA, without or with subsequent TdT tailing, respectively. Lanes 4 and 5 contained the products of reactions run with target DNA, probe oligo 32-161-2 (SEQ ID NO:87) and Invader™ oligo 32-161-1 (SEQ ID NO:88), without or with subsequent TdT tailing, respectively. Lanes 6 and 7 show the products of reactions containing target DNA, probe oligo 32-161-4 (SEQ ID NO:89) and Invader™ oligo 32-161-1 (SEQ ID NO:88), without or with subsequent TdT tailing, respectively. Lane M contains the marker ΦX174-*Hin*fI.

The reaction products in lanes 4 and 5 are the same as those seen in lanes 6 and 7, except that the absence of a 5' fluorescein on the probe prevents detection of the relased 5' product (indicated as "A" near the bottom of the gel) or the TdT extended 5' product (indicated as "B", near the top of the gel). The Cy3-labeled 3' portion of the cleaved probe is visible in all of these reactions (indicated as "C", just below the center of the gel).

To demonstrate detection of target-dependent Invader™-directed cleavage products on a solid support, the reactions from lanes 3 and 5 were tested on the Universal Genecomb® (Bio-Rad) which is a standard nitrocellulose matrix on a rigid nylon backing styled in a comb format, as depicted in Fig. 68. Following the manufacturer's protocol, with one modification: 10 μl of the Invader™-directed cleavage reactions were used instead the recommended 10% of a PCR. To capture the cleavage products, 2.5 pmols of the capture oligo 59-28-1 (SEQ ID NO:90) were spotted on each tooth. The capture and visualization steps were conducted according to the manufacturer's directions. The results are shown in Fig. 68.

In Fig. 68, teeth numbered 6 and 7 show the capture results of reactions performed without and with target DNA present. Tooth 8 shows the kit positive control.

The darkness of the spot seen on tooth 7, when compared to tooth 6, clearly indicates that products of Invader™-directed cleavage assays may be specifically detected on solid supports. While the Universal Genecomb® was used to demonstrate solid support capture in this instance, other support capture methods known to those skilled in the art would be equally suitable. For example, beads or the surfaces of reaction vessels may easily be coated with capture oligonucleotides so that they can then be used in this step. Alternatively, similar solid supports may easily be coated with streptavidin or antibodies for the capture of biotin-

or hapten-tagged products of the cleavage/tailing reaction. In any of these embodiments, the products may be appropriately visualized by detecting the resulting fluorescence, chemiluminescence, colorimetric changes, radioactive emissions, optical density change or any other distinguishable feature of the product.

EXAMPLE 31

Comparison Of The Effects Of Invasion Length and 5' Label Of The Probe On Invader™-Directed Cleavage By The Cleavase® A/G and *Pfu* FEN-1 Nucleases

To investigate the effect of the length of invasion as well as the effect of the type of dye on ability of *Pfu* FEN-1 and the Cleavase® A/G nuclease to cleave 5' arms, the following experiment was performed. Three probes of similar sequences labeled with either fluorescein, TET, or Cy3, were assembled in reactions with three Invader™ oligonucleotides which created overlapping target hybridization regions of eight, five, and three bases along the target nucleic acid, M13mp18.

The reactions were conducted as follows. All conditions were performed in duplicate. Enzyme mixes for *Pfu* FEN-1 and the Cleavase® A/G nuclease were assembled. Each 2 µl of the *Pfu* FEN-1 mix contained 100 ng of *Pfu* FEN-1 (prepared as described in Ex. 28) and 7.5 mM $MgCl_2$. Each 2 µl of the Cleavase® A/G mix contained 5.3 ng of the Cleavase® A/G nuclease and 4.0 mM $MnCl_2$. Six master mixes containing buffer, M13mp18, and Invader™ oligonucleotides were assembled. Each 7 µl of mixes 1-3 contained 1 fmol M13mp18, 10 pmoles Invader™ oligonucleotide (34-078-4 [SEQ ID NO:39], 24-181-2 [SEQ ID NO:91], or 24-181-1 [SEQ ID NO:92], in 10 mM MOPS (pH 7.5), 150 mM LiCl. Each 7 µl of mixes 4-6 contained 1 fmol of M13mp18, 10 pmoles of Invader™ oligonucleotide [34-078-4 (SEQ ID NO:39), 24-181-2 (SEQ ID NO:91), or 24-181-1 (SEQ ID NO:92)] in 10 mM Tris (pH 8.0). Mixtures 1-6 were then divided into three mixtures each, to which was added either the fluorescein-labeled probe (oligo 34-078-01; SEQ ID NO:86), the Cy3-labeled probe (oligo 43-20; SEQ ID NO:93) or the TET-labeled probe (oligo 90; SEQ ID NO:32 containing a 5' TET label). Each 7 µl of all mixtures contained 10 pmoles of corresponding probe. The DNA solutions described above were covered with 10 µl of ChillOut® evaporation barrier and brought to 68°C.

The reactions made from mixes 1-3 were started with 2 µl of the Cleavase® A/G nuclease mix, and the reactions made from mixes 4-6 were started with 2 µl of the *Pfu* FEN-1 mix. After 30 minutes at 68°C, the reactions were terminated by the addition of 8 µl of

95% formamide with 10 mM EDTA and 0.05% marker dyes. Samples were heated to 90°C for 1 minute immediately before electrophoresis through a 20% denaturing acrylamide gel (19:1 cross-linked) with 7 M urea, in a buffer containing 45 mM Tris-Borate (pH 8.3), 1.4 mM EDTA. The products of the cleavage reactions were visualized following electrophoresis by the use of a Hitachi FMBIO fluorescence imager. Results from the fluorescein-labeled probe are shown in Fig. 69, results from the Cy3-labeled probe in Fig. 70, and results from the TET-labeled probe in Fig. 71. In each of these Figures, the products of cleavage by Cleavase® A/G are shown in lanes 1-6 and the products of cleavage by *Pfu*FEN-1 are shown in lanes 7-12. In each in case the uncut material appears as a very dark band near the top of the gel, indicated by a "U" on the left. The products of cleavage directed by Invader™ oligonucleotides with 8, 5 or 3 bases of overlap (*i.e.*, the "X" region was 8, 5, or 3 nt long) are shown in the first, second and third pair of lanes in each set, respectively and the released labeled 5' ends from these reactions are indicated by the numbers 8, 5, and 3 on the left. Note that in the cleavage reactions shown in Fig. 70 the presence of the positively charged Cy3 dye causes the shorter products to migrate more slowly than the larger products. These products do not contain any additional positive charges (*e.g.*, amino modifications as used in Example 23), and thus still carry a net negative charge, and migrate towards the positive electrode in a standard electrophoresis run.

It can be seen from these data that the Cleavase® A/G and *Pfu* FEN-1 structure-specific nucleases respond differently to both dye identity and to the size of the piece to be cleaved from the probe. The *Pfu* FEN-1 nuclease showed much less variability in response to dye identity than did the Cleavase® A/G nuclease, showing that any dye wold be suitable for use with this enzyme. In contrast, the amount of cleavage catalyzed by the Cleavase® A/G nuclease varied substantially with dye identity. Use of the fluorescein dye gave results very close to those seen with the *Pfu* FEN-1 nuclease, while the use of either Cy3 or TET gave dramatically reduced signal when compared to the *Pfu* FEN-1 reactions. The one exception to this was in the cleavage of the 3 nt product carrying a TET dye (lanes 5 and 6, Fig. 71), in which the Cleavase® A/G nuclease gave cleavage at the same rate as the *Pfu* FEN-1 nuclease. These data indicate that, while Cleavase® A/G may be used to cleave probes labeled with these other dyes, the *Pfu* FEN-1 nuclease is a preferred nuclease for cleavage of Cy3- and TET-labeled probes.

EXAMPLE 32

Examination Of The Effects Of A 5' Positive Charge On The Rate Of Invasive Cleavage Using The Cleavase® A/G Or *Pfu* FEN-1 Nucleases

To investigate whether the positive charges on 5' end of probe oligonucleotides containing a positively charged adduct(s) (*i.e.*, charge reversal technology or CRT probes as described in Ex. 23 and 24 have an effect on the ability of the Cleavase® A/G or *Pfu* FEN-1 nucleases to cleave the 5' arm of the probe, the following experiment was performed.

Two probe oligonucleotides having the following sequences were utilized in Invader™ reactions: Probe 34-180-1: (N-Cy3)$T_{NH2}T_{NH2}$CCAGAGCCTAATTTGCC AGT(N-fluorescein)A, where N represents a spacer containing either the Cy3 or fluorescein group (SEQ ID NO:94) and Probe 34-180-2: 5'-(N-TET)TTCCAGAGCC TAATTTGCCAGT-(N-fluorescein)A, where N represents a spacer containing either the TET or fluorescein group (SEQ ID NO:95). Probe 34-180-1 has amino-modifiers on the two 5' end T residues and a Cy3 label on the 5' end, creating extra positive charges on the 5' end. Probe 34-180-2 has a TET label on the 5' end, with no extra positive charges. The fluorescein label on the 3' end of probe 34-180-1 enables the visualization of the 3' cleaved products and uncleaved probes together on an acrylamide gel run in the standard direction (*i.e.*, with the DNA migrating toward the positive electrode). The 5' cleaved product of probe 34-180-1 has a net positive charge and will not migrate in the same direction as the uncleaved probe, and is thus visualized by resolution on a gel run in the opposite direction (*i.e.*; with this DNA migrating toward the negative electrode).

The cleavage reactions were conducted as follows. All conditions were performed in duplicate. Enzyme mixes for the *Pfu* FEN-1 and Cleavase® A/G nucleases were assembled. Each 2 µl of the *Pfu* FEN-1 mix contained 100 ng of *Pfu* FEN-1 (prepared as described in Ex. 28) and 7.5 mM $MgCl_2$. Each 2 µl of the Cleavase® A/G nuclease mix contained 26.5 ng of Cleavase® A/G nuclease and 4.0 mM $MnCl_2$. Four master mixes containing buffer, M13mp18, and Invader™ oligonucleotides were assembled. Each 7 µl of mix 1 contained 5 fmol M13mp18, 10 pmoles Invader™ oligonucleotide 123 (SEQ ID NO:96) in 10 mM HEPES (pH 7.2). Each 7 µl of mix 2 contained 1 fmol M13mp18, 10 pmoles Invader™ oligonucleotide 123 in 10 mM HEPES (pH 7.2). Each 7 µl of mix 3 contained 5 fmol M13mp18, 10 pmoles Invader™ oligonucleotide 123 in 10 mM HEPES (pH 7.2), 250 mM KGlu. Each 7 µl of mix 4 contained 1 fmol M13mp18, 10 pmoles Invader™ oligonucleotide 123 in 10 mM HEPES (pH 7.2), 250 mM KGlu. For every 7 µl of each mix, 10 pmoles of

either probe 34-180-1 (SEQ ID NO:94) or probe 34-180-2 (SEQ IDNO:95) was added. The DNA solutions described above were covered with 10 µl of ChillOut® evaporation barrier and brought to 65°C. The reactions made from mixes 1-2 were started by the addition of 2 µl of the *Pfu* FEN-1 mix, and the reactions made from mixes 3-4 were started by the addition of 2 µl of the Cleavase® A/G nuclease mix. After 30 minutes at 65°C, the reactions were terminated by the addition of 8 µl of 95% formamide containing 10 mM EDTA. Samples were heated to 90°C for 1 minute immediately before electrophoresis through a 20% denaturing acrylamide gel (19:1 cross-linked) with 7 M urea, in a buffer containing 45 mM Tris-Borate (pH 8.3), 1.4 mM EDTA and a 20% native acrylamide gel (29:1 cross-linked) in a buffer containing 45 mM Tris-Borate (pH 8.3), 1.4 mM EDTA.

The products of the cleavage reactions were visualized following electrophoresis by the use of a Hitachi FMBIO fluorescence imager. The resulting images are shown in Fig. 72. Fig. 72A shows the denaturing gel which was run in the standard electrophoresis direction, and Fig. 72B shows the native gel which was run in the reverse direction. The reaction products produced by *Pfu* FEN-1 and Cleavase® A/G nucleases are shown in lanes 1-8 and 9-16, respectively. The products from the 5 fmol M13mp18 and 1 fmol M13mp18 reactions are shown in lanes 1-4, 9-12 (5 fmol) and 5-8, 13-16 (1 fmol). Probe 34-180-1 is in lanes 1-2, 5-6, 9-10, 13-14 and probe 34-180-2 is in lanes 3-4, 7-8, 11-12, 15-16.

The fluorescein-labeled 3' end fragments from all cleavage reactions are shown in Fig. 72A, indicated by a "3"' mark at the left. The 3 nt 5' TET-labeled products are not visible in this Figure, while the 5' Cy3-labeled products are shown in Fig. 72B.

The 3' end bands in Fig. 72A can be used to compare the rates of cleavage by the different enzymes in the presence of the different 5' end labels. It can be seen from this band that regardless of the amount of target nucleic acid present, both the *Pfu* FEN-1 and the Cleavase® A/G nucleases show more product from the 5' TET-labeled probe. With the *Pfu* FEN-1 nuclease this preference is modest, with only an approximately 25 to 40% increase in signal. In the case of the Cleavase® A/G nuclease, however, there is a strong preference for the 5' TET label. Therefore, although when the charge reversal method is used to resolve the products, a substantial amount of product is observed from the Cleavase® A/G nuclease-catalyzed reactions, the *Pfu* FEN-1 nuclease is a preferred enzyme for cleavage of Cy3-labeled probes.

EXAMPLE 33

The Use Of Universal Bases In The Detection Of Mismatches By Invader™ Directed Cleavage

The term "degenerate base" refers to a base on a nucleotide that does not hydrogen bond in a standard "Watson-Crick" fashion to a specific base complement (*i.e.*, A to T and G to C). For example, the inosine base can be made to pair via one or two hydrogen bonds to all of the natural bases (the "wobble" effect) and thus is called degenerate. Alternatively, a degenerate base may not pair at all; this type of base has been referred to as a "universal" base because it can be placed opposite any nucleotide in a duplex and, while it cannot contribute stability by base-pairing, it does not actively destabilize by crowding the opposite base. Duplexes using these universal bases are stabilized by stacking interactions only. Two examples of universal bases, 3-nitropyrrole and 5-nitroindole, are shown in Fig. 73. In hybridization, placement of a 3-nitropyrrole three bases from a mismatch position enhances the differential recognition of one base mismatches. The enhanced discrimination seems to come from the destabilizing effect of the unnatural base (*i.e.*, an altered T_m in close proximity to the mismatch). To test this same principle as a way of sensitively detecting mismatches using the Invader™-directed cleavage assay, Invader™ oligonucleotides were designed using the universal bases shown in Fig. 73, in the presence or absence of a natural mismatch. In these experiments, the use of single nitropyrrole bases or pairs of nitroindole bases that flank the site of the mismatch were examined.

The target, probe and Invader™ oligonucleotides used in these assays are shown in Fig. 74. A 43 nucleotide oligonucleotide (oligo 109; SEQ ID NO:97) was used as the target. The probe oligonucleotide (oligo 61; SEQ ID NO:50) releases a net positively charged labeled product upon cleavage. In Fig. 74, the Invader™ oligonucleotide is shown schematically above the target oligonucleotide as an arrow; the large arrowhead indicates the location of the mismatch between the Invader™ oligos and the target. Under the target oligonucleotide, the completely complementary, all natural (*i.e.*, no universal bases) Invader™ oligo (oligo 67; SEQ ID NO:51) and a composite of Invader™ oligos containing universal bases ("X") on either side of the mismatch ("M") are shown. The following Invader™ oligos were employed: oligo 114 (SEQ ID NO:98) which contains a single nt mismatch; oligo 115 (SEQ ID NO:99) which contains two 5-nitroindole bases and no mismatch; oligo 116 (SEQ ID NO:100) which contains two 5-nitroindole bases and a single nt mismatch; oligo 112 (SEQ ID NO:101) which contains one 3-nitropyrrole base and no mismatch; oligo 113 (SEQ

ID NO:102) which contains one 5-nitropyrrole base and a single nt mismatch; and oligo 67 (SEQ ID NO:51) which is completely complementary to the target.

The Invader™-directed cleavage reactions were carried out in 10 μl of 10 mM MOPS (pH 7.2), 100 mM KCl, containing 1 μM of the appropriate invading oligonucleotide (oligos 67, 112-116), 10 nM synthetic target 109, 1 μM Cy-3 labeled probe 61 and 2 units of Cleavase® DV (prepared as described in Ex. 27). The reactions were overlayed with Chill-Out® liquid wax, brought to the appropriate reaction temperature, 52°C, 55°C, or 58°C and initiated with the addition of 1 μl of 40 mM $MnCl_2$. Reactions were allowed to proceed for 1 hour and were stopped by the addition of 10 μl formamide. One fourth of the total volume of each reaction was loaded onto 20% non-denaturing polyacrylamide gels which were electrophoresed in the reverse direction. The products were visualized using an Hitachi FMBIO-100 fluorescent scanner using a 585 nm filter. The resulting images are shown in Figs. 75A-C. In each panel, lanes 1-6 contain reactions products from reactions using Invader™ oligo 67, 114, 115, 116, 112 and 113, respectively. Reactions run at 52°C, 55°C and 58°C are shown in Panels A, B and C, respectively.

These data show that two flanking 5-nitroindoles display a significantly greater differentiation then does the one 3-nitropyrrole system, or the all natural base hybridization, and this increased sensitivity is not temperature dependent. This demonstrates that the use of universal bases is a useful means of sensitively detecting single base mismatches between the target nucleic acid and the complex of detection oligonucleotides of the present invention.

EXAMPLE 34

Detection Of Point Mutations in The Human *Ras* Oncogene Using A Miniprobe

It is demonstrated herein that very short probes can be used for sensitive detection of target nucleic acid sequences (Ex. 37). In this Example, it is demonstrated that the short probes work very poorly when mismatched to the target, and thus can be used to distinguish a given nucleic acid sequence from a close relative with only a single base difference. To test this system synthetic human *ras* oncogene target sequences were created that varied from each other at one position. Oligonucleotide 166 (SEQ ID NO:103) provided the wild-type *ras* target sequence. Oligonucleotide 165 (SEQ ID NO:104) provided the mutant *ras* target sequence. The sequence of these oligonucleotides are shown in Fig. 76, and the site of the sequence variation in the site corresponding to codon 13 of the ras gene is indicated. The Invader™ oligonucleotide (oligo 162) has the sequence:

5'-$G_SC_ST_SC_SA_SA_SG_SG_SC_S$ACTCTTGCC TACGA-3' (SEQ ID NO:105), where the "S" indicates thiol linkages (*i.e.*, these are 2'-deoxynucleotide-5'-O-(1-thiomonophates)). The miniprobe (oligo 161) has the sequence: 5'-(N-Cy3) $T_{NH2}T_{NH2}$CACCAG-3' (SEQ ID NO:106) and is designed to detect the mutant *ras* target sequence (*i.e.*, it is completely complementary to oligo 165). The stacker oligonucleotide (oligo 164) has the sequence: 5'-$C_ST_SC_SC_SA_SA_SC_S$ T_SA_SCCACAAGTTTATATTCAG-3' (SEQ ID NO:107). A schematic showing the assembly of these oligonucleotides into a cleavage structure is depicted in Fig. 76.

Each cleavage reaction contained 100 nM of both the invading (oligo 162) and stacking (oligo 164) oligonucleotides, 10 μM Cy3-labeled probe (oligo 161) and 100 pM of either oligo 165 or oligo 166 (target DNA) in 10 μl of 10 mM HEPES (pH 7.2), 250 mM KGlu, 4 mM $MnCl_2$. The DNA mixtures were overlaid with mineral oil, heated to 90°C for 15 sec then brought to a reaction temperature of 47°, 50°, 53° or 56°C. Reactions were initiated by the addition of 1 μl of 100 ng/μl Pfu FEN-1. Reactions were allowed to proceed for 3 hours and stopped by the addition of 10 μl formamide. One fourth of the total volume od each reaction was loaded onto a 20% non-denaturing polyacrylamide gel which was electrophoresed in the reverse direction. The gel was scanned using an Hitachi FMBIO-100 fluorescent scanner fitted with a 585 nm filter, and the resulting image is shown in Fig. 77.

In Fig. 77, for each reaction temperature tested, the products from reactions containing either the mutant *ras* target sequence (oligo 165) or the wild-type (oligo 166) are shown.

These data demonstrate that the miniprobe can be used to sensitively discriminate between sequences that differ by a single nucleotide. The miniprobe was cleaved to produce a strong signal in the presence of the mutant target sequence, but little or no miniprobe was cleaved in the presence of the wild-type target sequence. Furthermore, the discrimination between closely related targets is effective over a temperature range of at least 10°C, which is a much broader range of temperature than can usually be tolerated when the selection is based on hybridization alone (*e.g.*, hybridization with ASOs). This suggests that the enzyme may be a factor in the discrimination, with the perfectly matched miniprobe being the preferred substrate when compared to the mismatched miniprobe. Thus, this system provides sensitive and specific detection of target nucleic acid sequences.

EXAMPLE 35

Effects of 3' End Identity On Site Of Cleavage Of A Model Oligonucleotide Structure

As described in the Examples above, structure-specific nucleases cleave near the junction between single-stranded and base-paired regions in a bifurcated duplex, usually about one base pair into the base-paired region. It was shown in Example 10 that thermostable 5' nucleases, including those of the present invention (*e.g.*, Cleavase® BN nuclease, Cleavase® A/G nuclease), have the ability to cleave a greater distance into the base paired region when provided with an upstream oligonucleotide bearing a 3' region that is homologous to a 5' region of the subject duplex, as shown in Fig. 26. It has also been determined that the 3' terminal nucleotide of the Invader™ oligonucleotide may be unpaired to the target nucleic acid, and still shift cleavage the same distance into the down stream duplex as when paired. It is shown in this Example that it is the base component of the nucleotide, not the sugar or phosphate, that is necessary to shift cleavage.

Figs. 78A and B shows a synthetic oligonucleotide which was designed to fold upon itself which consists of the following sequence: 5'-GTTCTCTGCTCTCTGGTC GCTGTCTCGCTTGTGAAACAAGCGAGACAGCGTGGTCTCTCG-3' (SEQ ID NO:29). This oligonucleotide is referred to as the "S-60 Hairpin." The 15 basepair hairpin formed by this oligonucleotide is further stabilized by a "tri-loop" sequence in the loop end (*i.e.*, three nucleotides form the loop portion of the hairpin) (Hiraro *et al.*, Nucleic Acids Res., 22(4): 576 [1994]). Fig. 78B shows the sequence of the P-15 oligonucleotide (SEQ ID NO:30) and the location of the region of complementarity shared by the P-15 and S-60 hairpin oligonucleotides. In addition to the P-15 oligonucleotide shown, cleavage was also tested in the presence of the P-14 oligonucleotide (SEQ ID NO:108) (P-14 is one base shorter on the 3' end as compared to P-15), the P-14 with an abasic sugar (P-14d; SEQ ID NO:109) and the P14 with an abasic sugar with a 3' phosphate (P-14dp; SEQ ID NO:110). A P-15 oligo with a 3' phosphate, P-15p (SEQ ID NO:111) was also examined. The black arrows shown in Fig. 78 indicate the sites of cleavage of the S-60 hairpin in the absence (top structure; A) or presence (bottom structure; B) of the P-15 oligonucleotide.

The S-60 hairpin molecule was labeled on its 5' end with fluorescein for subsequent detection. The S-60 hairpin was incubated in the presence of a thermostable 5' nuclease in the presence or the absence of the P-15 oligonucleotide. The presence of the full duplex which can be formed by the S-60 hairpin is demonstrated by cleavage with the Cleavase® BN

5' nuclease, in a primer-independent fashion (*i.e.*, in the absence of the P-15 oligonucleotide). The release of 18 and 19-nucleotide fragments from the 5' end of the S-60 hairpin molecule showed that the cleavage occurred near the junction between the single and double stranded regions when nothing is hybridized to the 3' arm of the S-60 hairpin (Fig. 27, lane 2).

The reactions shown in Fig. 78C were conducted in 10 μl 1X CFLP buffer with 1 mM $MnCl_2$ and 50 mM K-Glutamate, in the presence of 0.02 μM S-60, 0.5 μM Invader™ oligonucleotide and 0.01 ng per μl Cleavase® BN nuclease. Reactions were incubated at 40°C for 5 minutes and stopped by the addition of 8 μl of stop buffer (95% formamide, 20 mM EDTA, 0.02% methyl violet). Samples were heated to 75°C for 2 min immediately before electrophoresis through a 15% acrylamide gel (19:1 cross-linked), with 7 M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA. Gels were then analyzed with a FMBIO-100 Image Analyzer (Hitachi) equipped with 505 nm filter. The resulting image is shown in Fig. 78C.

In Fig. 78C lane 1 contains products from the no enzyme control; lane 2 contains products from a reaction run in the absence of an Invader™ oligo; lanes 3-6 contain products from reactions run the presence of the P-14d, P-14dp, P-15 and P-15p Invader™ oligos, respectively.

From the data shown in Fig. 78C, it can be seen that the use of the P-15 Invader™ oligonucleotide produces a shift in the cleavage site, while the P14 Invader™ oligonucleotide with either a ribose (P14d) or a phosphorylated ribose (P14dp) did not This indicates that the 15th residue of the Invader™ oligonucleotide must have the base group attached to promote the shift in cleavage. Interestingly, the addition of phosphate to the 3' end of the P15 oligonucleotide apparently reversed the shifting of cleavage site. The cleavage in this lane may in fact be cleavage in the absence of an Invader™ oligonucleotide as is seen in lane 2. In experiments with 5' dye-labeled Invader™ oligonucleotides with 3' phosphate groups these oligonucleotides have been severely retarded in gel migration, suggesting that either the enzyme or another constituent of the reaction (*e.g.*, BSA) is able to bind the 3' phosphate irrespective of the rest of the cleavage structure. If the Invader™ oligonucleotides are indeed being sequestered away from the cleavage structure, the resulting cleavage of the S-60 hairpin would occur in a "primer-independent' fashion, and would thus not be shifted.

In addition to the study cited above, the effects of other substituents on the 3' ends of the Invader™ oligonucleotides were investigated in the presence of several different enzymes, and in the presence of either Mn++ or Mg++. The effects of these 3' end modifications on

the generation of cleaved product are summarized in the following table. All of modifications were made during standard oligonucleotide synthesis by the use of controlled pore glass (CPG) synthesis columns with the listed chemical moiety provided on the support as the synthesis starting residue. All of these CPG materials were obtained from Glen Research Corp. (Sterling, VA).

Fig. 79 provides the structures for the 3' end substituents used in these experiments.

TABLE 4

Modification Studies At 3' End Of Invader™ Oligo

3'-End Modification	Extension By Terminal Transferase	Effect on Invader™ Rxn. (As Invader™) Enzyme:Condition - Effect
3' phosphate Glen part # 20-2900-42	no	A:5 - inhibits reaction, no detectable activity
3' acridine Glen part # 20-2973-42	yes, poorly	A:5 - decrease in activity, <10% B:5 - decrease in activity , < 10% B:4 - decrease in activity, < 10% C:1 -decrease in activity, <10% C:2 - decrease in activity, ~20% C:4 - decrease in activity, ~ 50% C:3 - decrease in activity, <5%
3' carboxylate Glen part # 20-4090-42	no	A:1 - decrease in activity , ~50% activity shift in cleavage site C:3 - reduces rate, <10% activity
3' nitropyrole Glen part # 20-2143-42	yes	A:5 - increase in activity, ~2X
3' nitroindole Glen part # 20-2144-42	yes	A:5 - decrease in activity, ~33% activity
3' arabinose Glen part # 10-4010-90	yes	A:5 - decrease in activity, ~50% activity
3'dideoxyUTP- flourescein	no	A:5 - decrease in activity, ~40% activity
3'-3' linkage Glen part # 20-0002-01	no	A:1 - equivalent cleavage activity shift in cleavage site C:3 - decrease in activity, ~25% activity
3' glyceryl Glen part # 20-2902-42	yes, very poorly	C:3 - decrease in activity, ~30% activity loss of specificity of cleavage (2 sites)
3' amino modifier C7 Glen part # 20-2957-42	yes	C:3 - decrease in activity, ~30% activity loss of specificity, multiple sites
3'deoxy, 2'OH Glen part # 20-2104-42	yes, very poorly	A:5 - decrease in activity, <20% activity B:5 - decrease in activity, <20% activity B:3 - decrease in activity, <20% activity C:1 - equivalent activity C:2 - equivalent activity C:4 - ? increase in activity C:3 - decrease in activity, ~40% activity

Enzymes:
- A) Cleavase® DV nuclease
- B) Cleavase® BN nuclease
- C) *Pfu* FEN-1

Condition:
- 1) 4mM $MnCl_2$, 150mM LiCl
- 2) 4mM $MnCl_2$, 50mM KCl
- 3) 7.5mM $MgCl_2$, no monovalent
- 4) 4mM $MgCl_2$, 50mM KCl
- 5) 10mM MgOAc, 50mM KCl

It can be seen from these data that many different modifications can be used on the 3' end of the Invader™ oligonucleotide without detriment. In various embodiments of the present invention, such 3' end modifications may be used to block, facilitate, or otherwise alter the hybridization characteristics of the Invader™ oligonucleotide, (*e.g.*, to increase discrimination against mismatches, or to increase tolerance of mismatches, or to tighten the association between the Invader™ oligonucleotide and the target nucleic acid). Some substituents may be used to alter the behavior of the enzyme in recognizing and cleaving within the assembled complex.

Altered 3' ends may also be used to prevent extension of the Invader™ oligonucleotide by either template-dependent or template-independent nucleic acid polymerases. The use of otherwise unmodified dideoxynucleotides (*i.e.*, without attached dyes or other moieties) are a particularly preferred means of blocking extension of Invader™ oligonucleotides, because they do not decrease cleavage activity, and they are absolutely unextendable.

EXAMPLE 36

Effect Of Probe Concentration, Temperature And A Stacker Oligonucleotide On The Cleavage Of Miniprobes By Invader™-Directed Cleavage

The stacker oligonucleotides employed to form cleavage structures may serve two purposes in the detection of a nucleic acid target using a miniprobe. The stacker oligonucleotide may help stabilize the interaction of the miniprobe with the target nucleic acid, leading to greater accumulation of cleaved probe. In addition, the presence of this oligo in the complex elongates the duplex downstream of the cleavage site, which may enhance the cleavage activity of some of the enzymes of the present invention. An example of different preferences for the length of this duplex by different structure-specific nucleases is seen in the comparison of the Cleavase® BN nuclease and the *Mja* FEN-1 nuclease cleavage of 8 bp and

12 bp duplex regions in Fig. 65. Increased affinity of the enzyme for the cleavage structure also results in increased accumulation of cleaved probe during reactions done for a set amount of time.

The amount of miniprobe binding to the target is also affected by the concentration of the miniprobe in the reaction mixture. Even when a miniprobe is only marginally likely to hybridize (*e.g.*, when the reaction is performed at temperatures in excess of the expected melting temperature of the probe/target duplex), the amount of probe on the target at any given time can be increased by using high concentrations of the miniprobe.

The need for a stacker oligonucleotide to enhance cleavage of the miniprobe was examined at both low and high probe concentrations. The reactions were carried out in 10 μl of 10 mM HEPES (pH 7.2), 250 mM KGlu, 4 mM $MnCl_2$, containing 100 nM of both the invading (oligo 135; SEQ ID NO:112) and stacking oligonucleotides (oligo 147; SEQ ID NO:113) and 100 pM ssM13 DNA. The reactions were overlayed with mineral oil, heated to 90°C for 15 sec then brought to the reaction temperature. Reactions were performed at 35°, 40°, 45°, 50°, 55°, 60°, and 65°C. The cleavage reactions were initiated by the addition of 1 μl of 100 ng/μl *Pfu* FEN-I and 1 μl of varying concentrations of Cy-3 labeled 142 miniprobe oligonucleotide (SEQ ID NO:114). Reactions were allowed to proceed for 1 hour and stopped by the addition of 10 μl formaldehyde. One fourth of the total volume of each reaction was loaded onto 20% non-denaturing polyacrylamide gels which were electrophoresed in the reverse direction. Gels were visualized using an Hitachi FMBIO-100 fluorescent scanner using a 585 nm filter. The fluorescence in each product band was measured and the graph shown in Fig. 80 was created using a Microsoft Excel spreadsheet.

The data summarized in Fig. 80 showed that the concentration of the miniprobe had a significant effect on the final measure of product, showing dramatic increases as the concentration was raised. Increases in the concentration of the miniprobe also shifted the optimum reaction temperature upward. It is known in the art that the concentration of the complementary strands in a hybridization will affect the apparent T_m of the duplex formed between them. More significantly to the methods and compositions of the present invention is the fact that the presence of the stacker oligonucleotide has a profound influence on the cleavage rate of the miniprobe at all probe concentrations. At each of the probe concentrations the presence of the stacker as much as doubled the signal from thc cleavage product. This demonstrated the utility of using the stacker oligonucleotide in combination with the miniprobes described herein.

EXAMPLE 37

The Presence of A Mismatch In The Invader™ Oligonucleotide Decreases The Cleavage Activity Of The Cleavase® A/G Nuclease

In any nucleic acid detection assay it is of additional benefit if the assay can be made to sensitively detect minor differences between related nucleic acids. In the following experiment, model cleavage substrates were used that were identical except for the presence or absence of a mismatch near the 3' end of the Invader™ oligonucleotide when hybridized to the model target nucleic acid. The effect of a mismatch in this region on the accumulation of cleaved probe was then assessed.

To demonstrate the effect of the presence of a mismatch in the Invader™ oligonucleotide on the ability of the Cleavase® A/G nuclease to cleave the probe oligonucleotide in an Invader™ assay the following experiment was conducted. Cleavage of the test oligonucleotide IT-2 (SEQ ID NO:115) in the presence of Invader™ oligonucleotides IT-1 (SEQ ID NO:116) and IT-1A4 (SEQ ID NO:117). Oligonucleotide IT-1 is fully complementary to the 3' arm of IT-2, whereas oligonucleotide IT-1A4 has a T->A substitution at position 4 from the 3' end that results in an A/A mismatch in the Invader™-target duplex. Both the matched and mismatched Invader™ oligonucleotides would be expected to hybridize at the temperature at which the following reaction was performed. Fig. 81 provides a schematic showing IT-1 annealed to the folded IT-2 structure and showing IT-1A4 annealed to the folded IT-2 structure.

The reactions were conducted as follows. Test oligonucleotide IT-2 (0.1 μM), labeled at the 5' end with fluorescein (Integrated DNA Technologies), was incubated with 0.26 ng/μl Cleavase® AG in 10 μl of CFLP® buffer with 4 mM $MgCl_2$, in the presence of 1 μM IT-1 or IT-1A4 at 40°C for 10 min; a no enzyme control was also run. Samples were overlaid with 15 μl Chill-Out® liquid wax to prevent evaporation. Reactions were stopped by addition of 4 μl stop buffer (95% formamide, 20 mM EDTA, 0.02% methyl violet). The cleavage products were separated on a 20% denaturing polyacrylamide gel and analyzed with the FMBIO-100 Image Analyzer (Hitachi) equipped with 505 nm filter. The resulting image is shown in Fig. 82.

In Fig. 82, lane 1 contains reaction products from the no enzyme control and shows the migration of the uncut IT-2 oligo; lanes 2-4 contain products from reactions containing no Invader™ oligo, the IT-1 Invader™ oligo and the IT-1A4 Invader™ oligo, respectively.

These data show that cleavage is markedly reduced by the presence of the mismatch, even under conditions in which the mismatch would not be expected to disrupt hybridization. This demonstrates that the Invader™ oligonucleotide binding region is one of the regions within the complex in which can be used for mismatch detection, as revealed by a drop in the cleavage rate.

EXAMPLE 38

Comparison Of The Activity Of The *Pfu* FEN-1 And *Mja* EN-1 Nucleases in the Invader™ Reaction

To compare the activity of the *Pfu* FEN-1 and the *Mja* FEN-1 nucleases in Invader™ reaction the following experiment was performed. A test oligonucleotide IT3 (SEQ ID NO:118) that forms an Invader™-Target hairpin structure and probe oligonucleotide PR1 (SEQ ID NO:119) labeled at the 5' end with fluorescein (Integrated DNA Technologies) were employed in Invader™ assays using either the *Pfu* FEN-1 or the *Mja* FEN-1 nucleases.

The assays were conducted as follows. *Pfu* FEN-1 (13 ng/μl) and *Mja* FEN-1 (10 ng/μl) (prepared as described in Ex. 28) were incubated with the IT3 (0.1 nM) and PR1 (2 and 5 μM) oligonucleotides in 10 μL CFLP® buffer, 4 mM $MgCl_2$, 20 mg/ml tRNA at 55°C for 41 min. Samples were overlaid with 15 μl Chill-Out® evaporation barrier to prevent evaporation. Reactions were stopped by addition of 70 μl stop buffer (95% formamide, 20 mM EDTA, 0.02% methyl violet). Reaction products (1 μl) were separated on a 20% denaturing polyacrylamide gel, visualized using a fluroimager and the bands corresponding to the probe and the product were quantitiated. The resulting image is shown in Fig. 83. In Fig. 83, the turnover rate per target per minute is shown below the image for each nuclease at each concentration of probe and target tested.

It was demonstrated in Example 32 that the use of the *Pfu* FEN-1 structure-specific nuclease in the Invader™-directed cleavage reaction resulted in a faster rate of product accumulation than did the use of the Cleavase® A/G. The data presented here demonstrates that the use of *Mja* FEN-1 nuclease with the fluorescein labeled probe further increases the amount of product generated by an average of about 50%, demonstrating that, in addition to the *Pfu* FEN-1 nuclease, the *Mja* FEN-1 nuclease is a preferred structure-specific nuclease for the detection of nucleic acid targets by the method of the present invention.

EXAMPLE 39

Detection Of RNA Target Nucleic Acids Using Miniprobe And Stacker Oligonucleotides

In addition to the detection of the M13 DNA target material described above, a miniprobe/stacker system was designed to detect the HCV-derived RNA sequences described in Example 19. A probe of intermediate length, either a long mid-range or a short standard probe, was also tested. The miniprobe used (oligo 42-168-1) has the sequence: 5'-TET-CCGGTCGTCCTGG-3' (SEQ ID NO:120), the stacker oligonucleotide used (oligo 32-085) with this miniprobe has the sequence: 5'-CAATTCCGGTGTACTACCGGTTCC-3' (SEQ ID NO:121). The slightly longer probe, used without a stacker (oligo 42-088), has the sequence: 5'-TET-CCGGTCGTCCTGGCAA-3' (SEQ ID NO:122). The Invader™ oligonucleotide used with both probes has the sequence: 5'-GTTTATCCAAGAAAGGACCCGGTC-3' (SEQ ID NO:47). The reactions included 50 fmole of target RNA, 10 pmole of the Invader™ oligonucleotide and 5 pmole of the miniprobe oligonucleotide in 10 μl of buffer containing 10 mM MES, pH 6.5 with 150 mM LiCl, 4 mM $MnCl_2$, 0.05% each Tween-20 and NP-40, and 39 units of RNAsin (Promega). When used, 10 pmoles of the stacker oligonucleotide was added. These components were combined, overlaid with Chillout® evaporation barrier, and warmed to 50°C; the reactions were started by the addition of 5 polymerase units of DNAPTth, to a final reaction volume of 10 μl. After 30 minutes at 50°C, reactions were stopped by the addition of 8 μl of 95% formamide, 10 mM EDTA and 0.02% methyl violet. The samples were heated to 90°C for 1 minute and 2.5 μl of each of these reactions were resolved by electrophoresis through a 20% denaturing polyacrylamide (19:1 cross link) with 7M urea in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA, and the labeled reaction products were visualized using the FMBIO-100 Image Analyzer (Hitachi). The resulting image is shown in Fig. 84.

In Fig. 84, lanes 1 and 2 show the products of reactions containing the HCV Invader™ oligonucleotide and the longer probe (oligo 42-088), without and with the target RNA present, respectively. Lanes 3, 4, and 5 show the products of reactions containing the Invader™ oligonucleotide and the shorter probe (oligo 42-168-1). Lane 3 is a control reaction without target RNA present, while lanes 4 and 5 have the target, but are without or with the stacker oligonucleotide, respectively.

Under these conditions the slightly longer (16 nt) probe oligonucleotide was cleaved quite easily without the help of a stacker oligonucleotide. In contrast, the shorter probe (13

nt) required the presence of the stacker oligonucleotide to produce detectable levels of cleavage. These data show that the miniprobe system of target detection by Invader™-directed cleavage is equally applicable to the detection of RNA and DNA targets. In addition, the comparison of the cleavage performance of longer and shorter probes in the absence of a stacker oligonucleotide give one example of the distinction between the performance of the miniprobe/stacker system and the performance of the mid-range and long probes in the detection of nucleic acid targets.

EXAMPLE 40

Effect Of An Unpaired 3' Tail On Transcription From A Complete (Un-Nicked) Promoter

In designing the method of transcription-based visualization of the products of Invader™-directed cleavage, it was first necessary to assess the effect of a 3' tail on the efficiency of transcription from a full length promoter. The duplexes tested in this Example are shown at the bottom of Fig. 93, and are shown schematically in Figs. 85A-C.

Transcription reactions were performed using the MEGAshortscript™ system from Ambion, Inc. (Austin, TX), in accordance with the manufacturer's instructions with the exception that a fluorescein labeled ribonucleotide was added. Each DNA sample was assembled in 4 μl of RNAse-free dH_2O. Reactions 1-3 each contained 10 pmole of the copy template oligo 150 (SEQ ID NO:123); reaction 2 contained 10 pmole of the promoter oligo 151 (SEQ ID NO:124); sample 3 contained 10 pmole of the 3' tailed promoter oligo 073-065 (SEQ ID NO:125); sample 4 had no added DNA. To each sample, 6 μl of a solution containing 1 μl of 10X Transcription Buffer, 7.5mM each rNTP, 0.125mM fluorescein-12-UTP (Boehringer) and 1 μl T7 MEGAshortscript™ Enzyme Mix was added. The samples were then incubated at 37°C for 1 hour. One microliter of RNase-free DNase 1 (2U/μl) was added to each sample and the samples were incubated an additional 15 minutes at 37°C. The reactions were then stopped by the addition of 10 μl of a solution of 95% formamide, 5mM Na_2EDTA, with loading dyes. All samples were heated to 95°C for 2 minutes and 4 μl of each sample were resolved by electrophoresis through a 20% denaturing acrylamide gel (19:1 cross-linked) with 7M urea, in a buffer containing 45mM Tris-Borate (pH 8.3), 1.4mM EDTA. The gel was analyzed with a FMBIO II fluorescence image analyzer, and the resulting image is shown in Fig. 93. The RNA produced by successful transcription appears near the middle of the panel, as indicated ("RNA").

Examination of the products of transcription shown in lanes 2 and 3 show that the presence of the 3' tail on the full-length promoter has an adverse affect on the efficiency of transcription, but does not shut it off completely. Because the objective of the transcription-based visualization assays of the present invention is to discriminate between uncleaved probe and the shorter products of the invasive cleavage assay (cut probe), these data indicate that production of a full-length promoter in the cleavage reaction would be difficult to resolve from the background created by transcription from promoters containing the uncleaved probe if no other oligonucleotides were included in the assay. Means of suppressing transcription from such a branched promoter are discussed in the Description of the Invention and discussed below in Ex. 43.

EXAMPLE 41

Examination Of The Influence Of The Position Of The Nick On The Efficiency Of Transcription From Partial And Complete Composite Bacteriophage T7 Promoters

In the Description of the Invention, the procedure for testing prospective promoter pieces for suitability in an invasive cleavage-linked assay is described. One aspect of the test is to examine the effect a chosen nick site has on the efficiency of transcription from the final composite promoter. In addition, the individual pieces of nicked promoter are tested for transcription activity in the presence of the full-length un-nicked strand. In this experiment, a comparison on these points is made between a composite promoter having a nick in the non-template strand between nucleotides -11 and -10 relative to the initiation site (+1), and a promoter having a nick on the same strand, but positioned between nucleotides -8 and -7. The Figure numbers for the schematic representations of the contents of each reaction are indicated below each lane (*e.g.*, 85A = Fig. 85A). The site where the nick would be in a fully assembled composite promoter using the reaction oligonucleotides is also indicated below each lane ("-11/-10" and "-8/-7").

Transcription reactions were performed using the MEGAshortscript™ system, in accordance with the manufacturer's instructions, but with the exception that a fluorescein labeled ribonucleotide was added. Each DNA sample was assembled in 4 μl of RNAse-free dH_2O. Reaction 1 had no added DNA. Reactions 2-9 each contained 10 pmole of the copy template oligo 150 (SEQ ID NO:123). Reactions 3 and 4 contained 10 pmole of the -11 "cut" probe (oligo 073-061-01; SEQ ID NO:127) or 20 pmole of the -10 partial promoter

oligo 073-061-02 (SEQ ID NO:130), respectively, and reaction 5 contained both. Reactions 6 and 7 contained either the 10 pmole of the -8 "cut" probe (oligo 073-062-01; SEQ ID NO:126) or 20 pmoles of the -7 partial promoter oligo 073-062-02 (SEQ ID NO:129), respectively, and reaction 8 contained them both. Reaction 9 contained 10 pmole of the intact promoter oligo 151 (SEQ ID NO:124).

The transcription reactions were initiated, incubated, terminated and the reaction products were resolved and imaged as described in Ex. 40. The resulting image is shown in Fig. 92. The reaction numbers correspond to the lane numbers above the image. The RNA created by successful transcription appears in the upper third of the image. Comparison to the positive control reaction (rxn. 9) shows that the full-length RNA produced by each of the composite promoters is the same size as that produced in the control reaction, indicated that transcription initiated at the same site in each reaction.

In Fig. 92, lanes 3, 4, and 5 compare transcription from the two species of partially assembled promoters (see schematics in Figs. 86A and B) and the fully assembled composite promoter (Fig. 88B) having a nick between nucleotides -11 and -10 relative to the start of transcription. It can be seen from these data that neither partial promoter (lanes 3 and 4) is able to support transcription of the copy template, but that the composite promoter (lane 5) with this nick site is strongly transcribed. Surprisingly, comparison to the control reaction (lane 9) shows that the presence of a nick at this site (-11/-10) actually enhances transcription. While not limiting the present invention to any particular mechanism, it is believed that the enhancement of transcription is a result of both suppressing the formation of the shorter abortive transcripts and by allowing greater accumulation of the full length product. This result is highly reproducible.

In Fig. 92, lanes 6, 7, and 8 compare transcription a similar set of partial and complete promoters in which the nick is shifted 3 residues closer to the transcription start site. Examination of lane 6 shows that the presence of 3 extra bases on the -8"cut" probe (compared to the -11 "cut" probe in lane 3) allow this partial promoter to initiate transcription. This indicates that the -8/-7 site would be a poor choice for use in this embodiment of the present invention.

This experiment demonstrates the process for determining the suitable placement of a nick within a promoter assembly to achieve the desired result. Similar tests can easily be designed for testing other nicks within the bacteriophage T7 promoter tested in this Example,

or for testing suitable nick placement in any desired phage, prokaryotic or eukaryotic promoter.

EXAMPLE 42

Detection Of The Products Of Invader™-Directed Cleavage Through Transcription From A Composite Promoter

The Examples described above indicate that a small oligonucleotide can be used to complete assembly of a composite T7 promoter, thereby enabling transcription from that promoter. Earlier Examples demonstrate that the invasive cleavage reaction can be used release specific small oligonucleotide products from longer probe oligonucleotides. In this Example, it is demonstrated that these two observations can be combined, and that the products of the invasive cleavage reaction can be used to complete a promoter and enable subsequent transcription. The schematic representations of the composite promoters tested in this Example are shown in Fig. 88.

Two invasive cleavage reactions were set up, one without (rxn. 1) and one with (rxn. 2) input target DNA. The reactions (1 and 2) comprised 10mM MOPS (pH 7.5), 0.05% Tween-20, 0.05% NP-40 and 20 pmoles probe oligo 073-067-01 (SEQ ID NO:132) and 10 pmoles Invader™ oligo 073-073-02 (SEQ ID NO:134) in a volume of 14 μl. Reaction 2 also included 100 fmoles M13mp18 ssDNA. The samples were placed at 60°C and 6 μl of a solution containing 20 ng of *Mja* FEN-1 and 40mM Mg_2Cl were added to each sample to start the reactions. The samples were incubated at 60°C for 30 minutes and stopped by the addition of 3 μl of 2.5M NaOAc, 83mM Na_2EDTA (pH 8.0). Each sample was transferred to a 1.5 ml microcentrifuge tube and then the DNAs were precipitated by the addition of 60 μl of chilled 100% ethanol, and were stored at -20°C for 20 minutes. The pellets were collected by microcentrifugation, washed once with 80% ethanol to remove excess salt, then dried under vacuum. The product of this invasive cleavage reaction is a 12 nt oligonucleotide having the sequence: 5'-CGAAATTAATAC-3' (SEQ ID NO:128), termed the -12 cut probe (same sequence as oligo 073-073-03).

For transcription, the dried samples were each dissolved in 4 μl of a solution containing 1 pmole copy template oligo 150 and 2 pmoles -11 partial promoter oligo 073-073-012 (SEQ ID NO:131). Control samples 3 and 4 each contained 1 pmole of the copy template oligo 150; sample 3 also contained 1 pmole probe oligo 073-067-01 (SEQ ID NO:132) and 2 pmoles -11 partial promoter oligo 073-073-012 (see structure 88A); sample 4

contained 1 pmole -12 "cut" probe oligo 073-073-03 (SEQ ID NO:128) and 2 pmoles -11 partial promoter oligo 073-073-012 (see structure 88B). These are the structures that would be expected to exist in the transcription reactions from the two invasive cleavage reactions described above.

The transcription reactions were initiated, incubated, terminated and the products were resolved and imaged as described in Ex. 40. The resulting image is shown in the right half of Fig. 89 (lanes 6-9). Samples 3 and 4 appear in lanes 6 and 7, respectively, and the reactions 1 and 2 from the invasive cleavage reaction products (indicated by the use of the lower case "i"), appear in lanes 8 and 9, respectively. The number of the Fig. showing the schematic representation of the expected promoter structure in each reaction is indicated above each lane, and the placement of the nick is also indicated. The uppercase letters indicate which structure in the particular Figure to examine for each reaction. The lowercase "i" above lanes 8 and 9 indicate that these transcriptions were derived from actual invasive cleavage reactions. These products are compared to the RNA produced in the control reaction in lane 5, the procedure for which is described in Ex. 44. The RNA created by successful transcription appears in the upper third of the panel (indicated by "RNA").

The reaction shown in lane 6 shows no transcription. This demonstrates that a nick between nucleotides -12 and -11 in the on-template strand of the T7 promoter eliminates transcription if the promoter is assembled from uncut probe such as the 3' end of the probe forms a branch within the promoter sequence. This is in contrast tot he results seen with the -11/-10 nick examined below. Further, the transcript apparent in lane 7 shows that an unbranched promoter with a nick at the same site (-12/-11) produces the correct RNA, with few abortive initiation products (see lanes 2 and 5 of Fig. 89, described in Ex. 44). The reactions in lanes 8 and 9 demonstrate that the same effect is observed when the invasive cleavage reaction is the sole source of the upstream piece (-12 cut probe) of the T7 promoter. It is worthy of note that the promoter that is transcribed in lane 8 is made complete by the presence of 1 pmole of a synthetic "cut" probe oligo, without any uncut probe in the mixture, while the promoter that is transcribed in lane 9 is completed by the product of an invasive cleavage reaction that had only 100 fmole of target DNA in it. This reaction also included the residual uncut probe (up to approx. 10 pmoles), which may compete for binding at the same site. Nonetheless, the transcriptions from the invasive cleavage reaction products are only slightly reduced in efficiency, and are just as free of background as is the "no target" sample (lane 8). This Example clearly demonstrates that the cleavage products from the

invasive cleavage reaction can be used in combination with a partial promoter oligo to promote the production of RNA, without background transcription generated by the presence of the uncut probe. This RNA product is clearly dependent on the presence of the target material in the invasive cleavage reaction.

EXAMPLE 43

Shutting Down Transcription From A "Leaky" Branched T7 Composite Promoter Through The Use Of A Downstream Partial Promoter Oligonucleotide Having A 5' Tail

The previous Example demonstrated that placement of a nick in the non-template strand of a bacteriophage T7 promoter between the -12 and -11 nucleotides, relative to the transcription start site, prevents transcription of the branched promoter while allowing transcription when the composite promoter is assembled using the cut probe. When the nick is placed in other locations in the T7 promoter, transcription may be initiated from either promoter, although it is usually less efficient from the branched promoter. This Example demonstrates that the addition of a 5' tail that can base pair to the uncut probe (Fig. 90A) to the downstream partial promoter piece effectively blocks transcription from that promoter, but does not prevent transcription when a cut probe completes the promoter (Fig. 90B).

Two invasive cleavage reactions were set up, one without (rxn. 7) and one with (rxn. 8) input target DNA. The reactions (7 and 8) comprised 10mM MOPS (pH 7.5), 0.05% Tween-20, 0.05% NP-40 and 20 pmoles probe oligo 073-067-01 (SEQ ID NO:132)and 10 pmoles Invader™ oligo 073-067-02 (SEQ ID NO:133) in a volume of 14 µl. Reaction 8 also included 100 fmoles M13mp18 ssDNA. The samples were placed at 60°C and 6 µl of a solution containing 20 ng of *Mja* FEN-1 and 40mM Mg_2Cl were added to each sample to start the reactions. The samples were incubated at 60°C for 30 minutes and then stopped by the addition of 3 µl of 2.5M NaOAc, 83mM Na_2EDTA (pH.8.0). Each sample was transferred to a 1.5 ml microcentrifuge tube and the DNAs were precipitated, washed and dried as described in Ex. 42. The product of this invasive cleavage reaction is 13 nt oligonucleotide sequence, 5'-CGAAATTAATACG-3' (SEQ ID NO:127), termed the -11 cut oligonucleotide sequence, 5'-CGAAATTAATACG-3' (SEQ ID NO:127), termed the -11 cut probe (same sequence as oligo 073-061-01 which is referred to as the -11 "cut" probe to indicate it was not generated in an invasive cleavage reaction).

In the transcription reactions, all of the DNAs were dissolved in 4 µl of RNase-free dH_2O. Sample 1 had no added DNA, samples 2-8 contained 1 pmole of the copy template

oligo 150 (SEQ ID NO:123). In addition, sample 3 contained 1 pmole of -11 "cut" probe oligo 073-061-01 (SEQ ID NO:127) and 2 pmoles of -10 partial promoter oligo 073-061-02 (SEQ ID NO:130), sample 4 contained 1 pmole of probe oligo 073-067-01 and 2 pmoles of -10 partial promoter oligo 073-061-02. Control sample 5 contained 1 pmole of probe oligo 073-067-01 and 2 pmoles of partial promoter w/5' tail oligo 073-074 (5'-TACTGACTCACTATAGGGTCTTCTATGGAG GTC-3' (SEQ ID NO:146) (*see* structure in Fig. 90A) and sample 6 contained 1 pmole of -11 "cut" probe oligo 073-061-01 and 2 pmoles of partial promoter w/5' tail oligo 073-074 (*see* structure in Fig. 90B). These are the structures (*i.e.*, 90A and 90B) that would be expected to exist in the transcription reactions from the two invasive cleavage reactions described above.

The dried samples 7 and 8 from the invasive cleavage (above) were each dissolved in 4 µl of dH_2O containing 1 pmole copy template oligo 150 and 2 pmoles partial promoter w/5' tail oligo 073-074. The transcription reactions were initiated, incubated, terminated and the reaction products were resolved and imaged as described in Ex. 40. The resulting image is shown in Fig. 91.

In Fig. 91 the lane numbers correspond to the sample numbers; the number of the Figure showing the schematic representation of the expected promoter structure in each reaction is indicated above each lane ("88" and "90"), and the placement of the nick is also indicated ("-11/-10"). The upper-case letters indicate which structure in the particular Figure to examine for each reaction. The lower case "i" above lanes 7 and 8 indicates that these transcriptions were derived from actual invasive cleavage reactions. The RNA created by successful transcription appears in the upper third of the panel, as indicated ("RNA").

The control reactions in lanes 1 and 2, having either no DNA or having the only the copy template, produced no RNA as expected. The product in lane 4 demonstrates that the branched T7 promoter with a nick in the non-template strand between nucleotides -11 and -10 can support transcription, albeit not as efficiently as the un-branched promoter with the nick at the same site (lane 3). Examination of lane 5 shows that the use of a partial promoter oligonucleotide with a short 5' tail that can basepair to the uncut probe as depicted in Fig. 90A, effectively suppresses this transcription but allows transcription when the probe does not have a 3' tail (lane 6; schematic Fig. 90B). The reactions in lanes 7 and 8 demonstrate that the same effect as observed when the invasive cleavage reaction is the sole source of the upstream piece (-11 cut probe, SEQ ID NO:127) of the T7 promoter. It is worthy of note that the promoter that is transcribed in sample 6 is made complete by the presence of 1 pmole

of a synthetic "cut probe", without any uncut probe in the mixture, while the promoter that is transcribed in sample 8 is completed by the product of an invasive cleavage reaction that had only 100 fmole of target DNA in it. This reaction also included the residual uncut probe (up to approximately 19 pmoles), which may compete for binding at the same site. Nonetheless, the transcriptions from the invasive cleavage reaction products are just as strong and just as free of background in the "no target" samples.

This Example clearly demonstrates that the cleavage products from the invasive cleavage reaction can be used in combination with a partial promoter oligonucleotide having a 5' tail to promote the production of RNA, without background transcription generated by the uncut probe. This RNA product is clearly dependent on the presence of the target material in the invasive cleavage reaction.

EXAMPLE 44

Creation Of A Complete Bacteriophage T7 Promoter By DNA Polymerase-Mediated Extension Of A Cut Probe Comprising A Partial T7 Promoter

As demonstrated in the Examples above, transcription cannot occur from the T7 promoter unless a complete promoter region is present. In the above Examples, a complete promoter containing a nick in one strand was created by annealing a cut probe generated from an invasive cleavage reaction to a copy template which was annealed to a partial promoter oligo. An alternative means of creating a complete promoter in a manner dependent upon detection of a target sequence in an invasive cleavage reaction is to anneal the cut probe to a copy template devoid of a partial promoter oligo. The 3'-OH present at the end of the annealed cut probe is then extended by a DNA polymerase to create a complete and un-nicked promoter which is transcription-competent.

In this Example, the promoter was made complete through the use of primer extension, rather that by the co-hybridization of another oligonucleotide. The reaction steps are diagrammed schematically in Fig. 87. Two invasive cleavage reactions were set up, one without (rxn. 1) and one with (rxn. 2) input target DNA. The reactions (1 and 2) comprised 10mM MOPS (pH 7.5), 0.05% Tween-20, 0.05% NP-40 and 20 pmoles probe oligo 073-067-01 (SEQ ID NO:132) and 10 pmoles Invader™ oligo 073-073-02 (SEQ ID NO:134) in a volume of 14 µl. Reaction 2 also included 100 fmoles M13mp18 ssDNA. The samples were placed at 60°C and 6 µl of a solution containing 20ng of *Mja* FEN-1 and 40mM Mg_2Cl were added to each sample to start the reactions. The samples were incubated at 60°C for 30

minutes and stopped by the addition of 3 μl of 2.5M NaOAc, 83mM Na_2EDTA (pH 8.0). Each sample was transferred to a 1.5 ml microcentrifuge tube and then the DNAs were precipitated, washed and dried as described in Ex. 42. The product of this invasive cleavage reaction is the 12 nt oligonucleotide sequence: 5'-CGAAATTAATAC-3' (SEQ ID NO:128), termed the -12 cut probe (same sequence as oligo 073-073-03 which is referred to as the -12 "cut" probe to indicate it was not generated in an invasive cleavage reaction).

To allow extension of these products using a template-dependent DNA polymerase, a 20 μl solution containing 20mM Tris-HCl (pH 8.5), 1.5mM Mg_2Cl, 50mM KCl, 0.05% Tween-20, 0.05% NP-40, 25μM each dNTP, 0.25 units *Taq* DNA polymerase (Boehringer) and 2 μM copy template oligo 150 (SEQ ID NO:123) was added to each of the dried cleavage samples. The samples were incubated at 30°C for 1 hr. The primer extension reactions were stopped by the addition of 3 μl of 2.5M NaOAc with 83mM Na_2EDTA (pH 8.0)/sample. Each sample was transferred to a 1.5 ml microcentrifuge tube and the DNAs were precipitated, washed and dried as described in Ex. 42.

Samples 1 and 2 were then dissolved in 4 μl RNase-free dH_2O. Samples 3, 4 and 5 are control reactions: sample 3 was 4 μl of RNase-free dH_2O without added DNA, sample 4 contained 1 pmole of the copy template oligo 150 (SEQ ID NO:123) in 4 μl of RNase-free dH_2O, and sample 5 contained 1 pmole of the same copy template and 1 pmole of the complete promoter oligo 151 (SEQ ID NO:124) in RNase-free dH_2O.

Transcription reactions were performed using the MEGAshortscript™ system, in accordance with the manufacturer's instructions, but with the addition of a fluorescein labeled ribonucleotide. To each sample, 6 μl of a solution containing 1 μl of 10X Transcription Buffer, 7.5mM each rNTP, 0.125mM fluorescein-12-UTP (Boehringer) and 1 μl T7 MEGAshortscript™ Enzyme Mix was added. The samples were incubated at 37°C for 1 hour. One μl of RNase-free DNase 1 (2U/μl) was added to each sample and they were incubated an additional 15 minutes at 37°C. The reactions were stopped by the addition of 10 μl of a solution of 95% formamide, 5mM NaEDTA, with loading dyes. All samples were heated to 95°C for 2 minutes and four μl of each sample were resolved by electrophoresis through a 20% denaturing acrylamide gel (19:1 cross-linked) with 7 M urea, in a buffer containing 45 mM Tris-Borate (pH 8.3), 1.4mM EDTA. The results were imaged using the Molecular Dynamics Fluoroimager 595, with excitation at 488 nm and, emission detected at 530 nm.

The resulting image is shown in lanes 1 through 5 of Fig. 89; the lane numbers correspond to the sample numbers. The Figure numbers corresponding to the schematic representations of the promoters transcribed in each reaction as indicated above the lanes. The RNA product from successful transcription appears in the upper third of the panel, as indicated ("RNA"). Unincorporated labeled nucleotide appears as a dense signal near the bottom ("NTPs"). Short transcription products caused by aborted initiation events (Milligan and Uhlenbeck, Methods Enzymol., 180:51 [1989]) appear as bands just above the free nucleotide in the lanes showing active transcription (*i.e.*, lanes 2 and 5).

It can clearly be seen from the data in lanes 1 and 2 that the transcription is dependent on the presence of the target material in the invasive cleavage reaction. It is shown elsewhere (see lane 3, Fig. 92) that the product of the cleavage reaction is not in itself sufficient to allow transcription from the copy template. Thus, the action of the DNA polymerase in extending the hybridized cut probe across the promoter is a necessary step in enabling the transcription in this embodiment. These data clearly demonstrate that both template-dependent extension by DNA polymerase, and extension followed by transcription are suitable methods of visualizing the products of the invasive cleavage assay. As discussed in the Description of the Invention, the products of thermal breakdown that possess 3' terminal phosphates would not be extended, and would thus be precluded from contributing to background transcription.

EXAMPLE 45

Test For The Dependence Of An Enzyme On The Presence Of An Upstream Oligonucleotide

When choosing a structure-specific nuclease for use in a sequential invasive cleavage reaction it is preferable that the enzyme have little ability to cleave a probe 1) in the absence of an upstream oligonucleotide, and 2) in the absence of overlap between the upstream oligonucleotide and the downstream labeled probe oligonucleotide. Figs. 99a-e depicts the several structures that can be used to examine the activity of an enzyme that is confronted with each of these types of structures. The structure **a** (Fig. 99a) shows the alignment of a probe oligonucleotide with a target site on bacteriophage M13 DNA (M13 sequences shown in Fig. 99 are provided in SEQ ID NO:163) in the absence of an upstream oligonucleotide. Structure **b** (Fig. 99b) is provided with an upstream oligonucleotide that does not contain a region of overlap with the labeled probe (the label is indicated by the star). In structures **c**, **d**

and **e** (Figs. 99c-e) the upstream oligonucleotides have overlaps of 1, 3 or 5 nucleotides, respectively, with the downstream probe oligonucleotide and each of these structures represents a suitable invasive cleavage structure. The enzyme *Pfu* FEN-1 was tested for activity on each of these structures and all reactions were performed in duplicate.

Each reaction comprised 1 μM 5' TET labeled probe oligonucleotide 89-15-1 (SEQ ID NO:152), 50 nM upstream oligonucleotide (either oligo 81-69-2 [SEQ ID NO:153], oligo 81-69-3 [SEQ ID NO:154], oligo 81-69-4 [SEQ ID NO:155], oligo 81-69-5 [SEQ ID NO:156], or no upstream oligonucleotide), 1 fmol M13 target DNA, 10 mg/ml tRNA and 10 ng of *Pfu* FEN-1 in 10 μl of 10 mM MOPS (pH 7.5), 7.5 mM $MgCl_2$ with 0.05% each of Tween 20 and Nonidet P-40.

All of the components except the enzyme and the $MgCl_2$ were assembled in a final volume of 8 μl and were overlaid with 10 μl of Chill-Out™ liquid wax. The samples were heated to the reaction temperature of 69°C. The reactions were started by the addition of the *Pfu* FEN-1 and $MgCl_2$, in a 2 μl volume. After incubation at 69°C for 30 minutes, the reactions were stopped with 10 μl of 95% formamide, 10 mM EDTA, 0.02% methyl violet. Samples were heated to 90°C for 1 min immediately before electrophoresis through a 20% denaturing acrylamide gel (19:1 cross-linked), with 7 M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA. Gels were then analyzed with a FMBIO-100 Hitachi FMBIO fluorescence imager. The resulting image is displayed in Fig. 100.

In Fig. 100, lanes labeled "a" contain the products generated from reactions conducted without an upstream oligonucleotide (structure **a**), lanes labelled "b" contain an upstream oligonucleotide which does not invade the probe/target duplex (structure **b**). Lanes labelled "c", "d" and "e" contain the products generated from reactions conducted using an upstream oligonucleotide that invades the probe/target duplex by 1, 3 or 5 bases, respectively. The size (in nucleotides) of the uncleaved probe and the cleavage products is indicated to the left of the image in Fig. 100.

As shown in Fig. 100, cleavage of the probe was not detectable when structures **a** and **b** were utilized. In contrast, cleavage products were generated when invasive cleavage structures were utilized (structures **c-e**). These data show that the *Pfu* FEN-1 enzyme requires an overlapping upstream oligonucleotide for specific cleavage of the probe.

Any enzyme may be examined for its suitability for use in a sequential invasive cleavage reaction by examining the ability of the test enzyme to cleave structures **a-e** (it is understood by those in the art that the specific oligonucleotide sequences shown in Figs. 99a-

e need not be employed in the test reactions; these structures are merely illustrative of suitable test structures). Desirable enzymes display little or no cleavage of structures **a** and **b** and display specific cleavage of structures **c-e** (*i.e.*, they generate cleavage products of the size expected from the degree of overlap between the two oligonucleotides employed to form the invasive cleavage structure).

EXAMPLE 46

Use Of The Products Of A First Invasive Cleavage Reaction To Enable A Second Invasive Cleavage Reaction With A Net Gain In Sensitivity

As discussed in the Description of The Invention above, the detection sensitivity of the invasive cleavage reaction can be increased by the performing a second round of invasive cleavage using the products of the first reaction to complete the cleavage structure in the second reaction (shown schematically in Fig. 96). In this Example, the use of a probe which when cleaved in a first invasive cleavage reaction forms an integrated Invader™ oligo and target molecule for use in a second invasive cleavage reaction is illustrated (shown schematically in Fig. 97).

A first probe was designed to contain some internal complementarity so that when cleaved in a first invasive cleavage reaction the product ("Cut Probe 1") could form a target strand comprising an integral Invader™ oligonucleotide, as depicted in Fig. 97. A second probe was provided in the reaction that would be cleaved at the intended site when hybridized to the newly formed target/Invader™ (Fig. 97). To demonstrate the gain in signal due to the performance of sequential invasive cleavages, a standard invasive cleavage assay, as described above, was performed in parallel.

All reactions were performed in duplicate. Each standard (*i.e.*, non-sequential) invasive cleavage reaction comprised 1 μM 5' fluorescein-labeled probe oligo 073-182 (5' Fl-AGAAAGGAAGGGAAGAAAGCGAA-3'; SEQ ID NO:157), 10 nM upstream oligo 81-69-4 (5'-CTTGACGGGGAAAGCCGGCGAACGTGGCGA-3'; SEQ ID NO:155), 10 to 100 attomoles of M13 target DNA, 10 mg/ml tRNA and 10 ng of *Pfu* FEN-1 in 10 μl of 10 mM MOPS (pH 7.5), 8 mM $MgCl_2$ with 0.05% each of Tween 20 and Nonidet P-40. All of the components except the enzyme and the $MgCl_2$ were assembled in a volume of 7 μl and were overlaid with 10 μl of Chill-Out™ liquid wax. The samples were heated to the reaction temperature of 62°C. The reactions were started by the addition of the *Pfu* FEN-1 and

$MgCl_2$, in a 2 μl volume. After incubation at 62°C for 30 minutes, the reactions were stopped with 10 μl of 95% formamide, 10 mM EDTA, 0.02% methyl violet.

Each sequential invasive cleavage reaction comprised 1 μM 5' fluorescein-labeled oligonucleotide 073-191 (the first probe or "Probe 1", 5' Fl-TGGAGGTCAAAACATCG ATAAGTCGAAGAAAGGAAGGGAAGAAAT-3'; SEQ ID NO:158), 10 nM upstream oligonucleotide 81-69-4 (5'- CTTGACGGGGAAA GCCGGCGAACGTGGCGA-3'; SEQ ID NO:155), 1 μM of 5' fluorescein labeled oligonucleotide 106-32 (the second probe or "Probe 2", 5' Fl-TGTTTTGACCT CCA-3'; SEQ ID NO:159), 1 to 100 amol of M13 target DNA, 10 mg/ml tRNA and 10 ng of *Pfu* FEN-1 in 10 μl of 10 mM MOPS (pH 7.5), 8 mM $MgCl_2$ with 0.05% each of Tween 20 and Nonidet P-40. All of the components except the enzyme and the $MgCl_2$ were assembled in a volume of 8 μl and were overlaid with 10 μl of Chill-Out™ liquid wax. The samples were heated to the reaction temperature of 62°C (this temperature is the optimum temperature for annealing of Probe 1 to the first target). The reactions were started by the addition of *Pfu* FEN-1 and $MgCl_2$, in a 2 μl volume. After incubation at 62°C for 15 minutes, the temperature was lowered to 58°C (this temperature is the optimum temperature for annealing of Probe 2 to the second target) and the samples were incubated for another 15 min. Reactions were stopped by the addition of 10 μl of 95% formamide, 20 mM EDTA, 0.02% methyl violet.

Samples from both the standard and the sequential invasive cleavage reactions were heated to 90° C for 1 min immediately before electrophoresis through a 20% denaturing acrylamide gel (19:1 cross-linked), with 7 M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA. The gel was then analyzed with a Molecular Dynamics FluorImager 595. The resulting image is displayed in Fig. 101a. A graph showing measure of fluorescence intensity for each of the product bands is shown in Fig. 101b.

In Fig. 101a, lanes 1-5 contain the products generated in standard invasive cleavage reactions that contained either no target (lane 1), 10 amol of target (lanes 2 and 3) or with 100 amol of target (lanes 4 and 5). The uncleaved probe is seen as a dark band in each lane about half way down the panel and the cleavage products appear as a smaller black band near the bottom of the panel, the position of the cleavage product is indicated by an arrow head to the left of Fig. 101a. The gray ladder of bands seen in lanes 1-5 is due to the thermal degradation of the probe as discussed above and is not related to the presence or absence of the target DNA. The remaining lanes display products generated in sequential invasive cleavage reactions that contained 1 amol of target (lanes 6 and 7), 10 amol of target (lanes 8

and 9) and 100 amol of target (lanes 10 and 11). The uncleaved first probe (Probe 1; labeled "1 uncut") is seen near the top of the panel, while the cleaved first probe is indicated as "1: cut". Similarly, the uncleaved and cleaved second probe are indicated as "2: uncut" and 2: cut," respectively.

The graph shown in Fig. 101b compares the amount of product generated from the standard reaction ("Series 1") to the amount of product generated from the second step of the sequential reaction ("Series 2"). The level of background fluorescence measured from a reaction that lacked target DNA was subtracted from each measurement. It can be seen from the table located below the graph that the signal from the standard invasive cleavage assay that contained 100 attomoles of target DNA was nearly identical to the signal from the sequential invasive cleavage assay in which 1 attomole of target was present, indicating that the inclusion of a second cleavage structure increases the sensitivity of the assay 100 to 200-fold. This boost in signal allows easy detection of target nucleic acids at the sub attomole level using the sequential invasive cleavage assay, while the standard assay, when performed using this enzyme for only 30 minutes, does not generate detectable product in the presence of 10 attomoles of target.

When the amount of target was decreased by 10 or 100 fold in the sequential invasive cleavage assay, the intensity of the signal was decreased by the same proportion. This indicates that the quantitative capability of the invasive cleavage assay is retained even when reactions are performed in series, thus providing a nucleic acid detection method that is both sensitive and quantitative.

While in this Example, the two probes used had different optimal hybridization temperatures (*i.e.*, the temperature empiracally determined to give the greatest turnover rate in the given reaction conditions), the probes may also be selected (*i.e.*, designed) to have the same optimal hybridization temperature so that a temperature shift during incubation is not necessary.

EXAMPLE 47

The Products Of A Completed Sequential Invasive Cleavage Reaction Cannot Cross Contaminate Subsequent Similar Reactions

As discussed in the Description of the Invention, the serial nature of the multiple invasive cleavage events that occur in the sequential invasive cleavage reaction, in contrast to the reciprocating nature of the polymerase chain reaction and similar doubling assays, means

that the sequential invasive cleavage reaction is not subject to contamination by the products of like reactions because the products of the first cleavage reaction do not participate in the generation of new signal in the second cleavage reaction. If a large amount of a completed reaction were to be added to a newly assembled reaction, the background that would be produced would come from the amount of target that was also carried in, combined with the amount of already-cleaved probe that was carried in. In this Example, it is demonstrated that a very large portion of a primary reaction must be introduced into the secondary reaction to create significant signal.

A first or primary sequential invasive cleavage reaction was performed as described above using 100 amol of target DNA. A second set of 5 reactions were assembled as described in Ex. 46 with the exception that portions of the first reaction were introduced and no additional target DNA was included. These secondary reactions were initiated and incubated as described above, and included 0, 0.01, 0.1, 1, or 10% of the first reaction material. A control reaction including 100 amol of target was included in the second set also. The reactions were stopped, resolved by electrophoresis and visualized as described above, and the resulting image is displayed in Fig. 102. The primary probe, uncut second probe and the cut 2nd probe are indicated on the left as "1: cut", 2: uncut" and 2: cut", respectively.

In Fig. 102, lane 1 shows the results of the first reaction with the accumulated product at the bottom of the panel, and lane 2 show a 1:10 dilution of the same reaction, to demonstrate the level of signal that could be expected from that level of contamination, without further amplification. Lanes 3 through 7 show the results of the secondary cleavage reactions that contained 0, 10, 1, 0.1 or 0.01% of the first reaction material added as contaminant, respectively and lane 8 shows a control reaction that had 100 amol of target DNA added to verify the activity of the system in the secondary reaction. The signal level in lane 4 is as would be expected when 10% of the pre-cleaved material is transferred (as in lane 2) and 10% of the transferred target material from the lane 1 reaction is allowed to further amplify. At all levels of further dilution the signal is not readily distinguished from background. These data demonstrate that while a large-scale transfer from one reaction to another may be detectable, cross contamination by the minute quantities that would be expected from aerosol or from equipment contamination would not be easily mistaken for a false positive result. These data also demonstrate that when the products of one reaction are deliberately carried over into a fresh sample, these products do not participate in the new

reaction, and thus do not affect the level of target-dependent signal that may be generated in that reaction.

EXAMPLE 48

Detection of Human Cytomegalovirus Viral DNA by Invasive Cleavage

The previous Example demonstrates the ability of the invasive cleavage reaction to detect minute quantities of viral DNA in the presence of human genomic DNA. In this Example, the probe and Invader™ oligonucleotides were designed to target the 3104-3061 region of the major immediate early gene of human cytomegalovirus (HCMV) as shown in Fig. 103. In Fig. 103, the Invader™ oligo (89-44; SEQ ID NO:160) and the fluorescein (Fl)-labeled probe oligo (89-76; SEQ ID NO:161) are shown annealed along a region of the HCMV genome corresponding to nucleotides 3057-3110 of the viral DNA (SEQ ID NO:162). The probe used in this Example is a poly-pyrimidine probe and as shown herein the use of a poly-pyrimidine probe reduces background signal generated by the thermal breakage of probe oligos.

The genomic viral DNA was purchased from Advanced Biotechnologies, Incorporated (Columbia, MD). The DNA was estimated (but not certified) by personnel at Advanced Biotechnologies to be at a concentration of 170 amol (1×10^8 copies) per microliter. The reactions were performed in quadruplicate. Each reaction comprised 1 μM 5' fluorescein labeled probe oligonucleotide 89-76 (SEQ ID NO:161), 100 nM Invader™ oligonucleotide 89-44 (SEQ ID NO:160), 1 ng/ml human genomic DNA, and one of five concentrations of target HCMV DNA in the amounts indicated above each lane in Fig. 104, and 10 ng of *Pfu* FEN-1 in 10 μl of 10 mM MOPS (pH 7.5), 6 mM $MgCl_2$ with 0.05% each of Tween 20 and Nonidet P-40. All of the components except the labeled probe, enzyme and $MgCl_2$ were assembled in a final volume of 7 μl and were overlaid with 10 μl of Chill-Out™ liquid wax. The samples were heated to 95°C for 5 min, then reduced to 62°C. The reactions were started by the addition of probe, *Pfu* FEN-1 and $MgCl_2$, in a 3 μl volume. After incubation at 62°C for 60 minutes, the reactions were stopped with 10 μl of 95% formamide, 10 mM EDTA, 0.02% methyl violet. Samples were heated to 90° C for 1 min immediately before electrophoresis through a 20% acrylamide gel (19:1 cross-linked), with 7 M urea, in a buffer of 45 mM Tris-Borate, pH 8.3, 1.4 mM EDTA. Gels were then analyzed with a Molecular Dynamics FluorImager 595.

The resulting image is displayed in Fig. 104. The replicate reactions were run in groups of four lanes with the target HCMV DNA content of the reactions indicated above each set of lanes (0-170 amol). The uncleaved probe is seen in the upper third of the panel ("Uncut 89-76") while the cleavage products are seen in the lower two-thirds of the panel ("Cut 89-76"). It can be seen that the intensity of the accumulated cleavage product is proportional to the amount of the target DNA in the reaction. Furthermore, it can be clearly seen in reactions that did not contain target DNA ("no target") that the probe is not cleaved, even in a background of human genomic DNA. While 10 ng of human genomic DNA was included in each of the reactions shown in Fig. 104, inclusion of genomic DNA up to 200 ng has slight impact on the amount of product accumulated. The data did not suggest that 200 ng per 10 μl of reaction mixture represented the maximum amount of genomic DNA that could be tolerated without a significant reduction in signal accumulation. For reference, this amount of DNA exceeds what might be found in 0.2 ml of urine (a commonly tested amount for HCMV in neonates) and is equivalent to the amount that would be found in about 5 μl of whole blood.

These results demonstrate that the standard (*i.e.*, non-sequential) invasive cleavage reaction is a sensitive, specific and reproducible means of detecting viral DNA. It can also be seen from these data that the use of a poly-pyrimidine probe reduces the background from thermal breakage of the probe, as discussed in Example 22. Detection of 1.7 amol of target is roughly equivalent to detection of 10^6 copies of the virus. This is equivalent to the number of viral genomes that might be found in 0.2 mls of urine from a congenitally infected neonate (10^2 to 10^6 genome equivalents per 0.2 mls; Stagno *et al.*, J. Infect. Dis., 132:568 [1975]). Use of the sequential invasive cleavage assay would permit detection of even fewer viral DNA molecules, facilitating detection in blood (10^1 to 10^5 viral particles per ml; Pector *et al.*, J. Clin. Microbiol., 30:2359 [1992]) which carries a much larger amount of heterologous DNA.

From the above it is clear that the invention provides reagents and methods to permit the detection and characterization of nucleic acid sequences and variations in nucleic acid sequences. The Invader™-directed cleavage reaction and the sequential Invader™-directed cleavage reaction of the present invention provide ideal direct detection methods that combine the advantages of the direct detection assays (*e.g.*, easy quantification and minimal risk of carry-over contamination) with the specificity provided by a dual or tri oligonucleotide hybridization assay.

As indicated in the Description of the Invention, the use of sequential invasive cleavage reactions can present the problem of residual uncut first, or primary, probe interacting with the secondary target, and either competing with the cut probe for binding, or creating background through low level cleavage of the resulting structure. This is shown diagramatically in Figs. 105 and 106. In Fig. 105, the reaction depicted makes use of the cleavage product from the first cleavage structure to form an Invader™ oligonucleotide for a second cleavage reaction. The structure formed between the secondary target, the secondary probe and the uncut primary probe is depicted in Fig. 105, as the right hand structure shown in step 2a. This structure is recognized and cleaved by the 5' nucleases, albeit very inefficiently (*i.e.*, at less than about 1% in most reaction conditions). Nonetheless, the resulting product is indistinguishable from the specific product, and thus may lead to a false positive result. The same effect can occur when the cleaved primary probe creates and integrated Invader™/target (IT) molecule, as described in Example 46; the formation of the undesirable complex is depicted schematically in Fig. 106, as the right hand structure shown in step 2a.

The improvements provided by the inclusion of Arrestor™ oligonucleotides of various compositions in each of these types of sequential Invader™ assays are demonstrated in the following Examples. These Arrestors™ are configured to bind the residual uncut probe from the first cleavage reaction in the series, thereby increasing the efficacy of and reducing the non-specific background in the subsequent reaction(s).

EXAMPLE 49

"Arrestor™" Oligonucleotides Improve Sensitivity of Multiple Sequential Invasive Cleavage Assays

In this Example, the effect of including an Arrestor™ oligonucleotide on the generation of signal using the IT probe system depicted in Figs. 97 and 106 is demonstrated. The Arrestor™ oligonucleotide hybridizes to the primary probe, mainly in the portion that recognizes the target nucleic acid during the first cleavage reaction. In addition to examining the effects of adding an Arrestor™, the effects of using Arrestor™ oligonucleotides that extended in complementarity different distances into the region of the primary probe that composes the secondary IT structure were also investigated. These effects were compared in reactions that included the target DNA over a range of concentrations, or that lacked target

DNA, in order to demonstrate the level of nonspecific (*i.e.*, not related to target nucleic acid) background in each set of reaction conditions.

The target DNA for these reactions was a fragment that comprised the full length of the hepatitis B genome from strain of serotype adw. This material was created using the polymerase chain reaction from plasmid pAM6 (ATCC #45020D). The PCRs were conducted using a vector-based forward primer, oligo # 156-022-001 (5'-ggcgaccacacccgtcctgt-3'; SEQ ID NO:168) and a reverse primer, oligo #156-022-02 (5'-ccacgatgcgtccggcgtag-3'; SEQ ID NO:169) to amplify the full length of the HBV insert, an amplicon of about 3.2kb. The cycling conditions included a denaturation of the plasmid at 95°C for 5 minutes, followed by 30 cycles of 95°C, 30 seconds; 60°C, 40 seconds; and 72°C, 4 minutes. This was followed by a final extension at 72°C for 10 minutes. The resulting amplicon, termed pAM6#2, was adjusted to 2 M NH_4OAc, and collected by precipitation wiht isopropanol. After drying *in vacuo*, DNA was dissolved in 10 mM Tris pH .0, 0.1 mM EDTA. The concentration was determined by OD_{200} measurement, and by Invader™ assay with comparison to a standard of known concentration.

The Invader™ reactions were conducted as follows. Five master mixes, termed "A," "B," "C," "D," and "E," were assembled; all mixes contained 12.5 mM MOPS, pH 7.5, 500 fmoles primary Invader™ oligo #218-55-05 (SEQ ID NO:171), 10 ng human genomic DNA (Novagen) and 30 ng AfuFEN1 enzyme, for every 8 μl of mix. Mix A contained no added HBV genomic amplicon DNA; mix B contained 600 molecules of HBV genomic amplicon DNA pAM6 #2; mix C contained 6,000 molecules pAM6 #2; mix D contained 60,000 molecules pAM6 #2; and mix E contained 600,000 molecules pAM6 #2. The mixes were aliquotted to the reaction tubes, 8 μl/tube: mix A to tubes 1, 2, 11, 12, 21 and 22; mix B to tubes 3, 4, 13, 14, 23 and 24; mix C to tubes 5, 6, 15, 16, 25 and 26; mix D to tubes 7, 8, 17, 18, 27 and 28; and mix E to tubes 9, 10, 19, 20, 29 and 30. The samples were incubated at 95°C for 4 minutes to denature the HBV genomic amplicon DNA. The reactions were then cooled to 67°C, and 2 μl of a mix containing 37.5 mM $MgCl_2$ and 2.5 pmoles 218-95-06 (SEQ ID NO:183) for every 2 μl, was added to each sample. The samples were incubated at 67°C for 60 minutes. Three secondary reaction master mixes were prepared, all mixes contained 10 pmoles of secondary probe oligonucleotide #228-48-04 (SEQ ID NO:173) for every 2 μl of mix. Mix 2A contained no additional oligonucleotide, mix 2B contained 5pmoles "Arrestor™" oligo # 218-95-03 (SEQ ID NO:184) and mix 2C contained 5 pmoles of "Arrestor™" oligo # 218-95-01 (SEQ ID NO:174). After the 60 minute incubation at

67°C (the primary reaction described above), 2 μl of the secondary reaction mix was added to each sample: Mix 2A was added to samples #1-10; Mix 2B was added to samples #11-20; and Mix 2C was added to samples #21-30. The temperature was adjusted to 52°C and the samples were incubated for 30 minutes at 52°C. The reactions were then stopped by the addition of 10 μl of a solution of 95% formamide, 5 mM EDTA and 0.02% crystal violet. All samples were heated to 95°C for 2 minutes, and 4 μl of each sample were resolved by electrophoresis through 20% denaturing acrylamide gel (19:1 cross-linked) with 7 M urea, in a buffer containing 45 mM Tris-Borate (pH8.3) and 1.4 mM EDTA. The results were imaged using the Molecular Dynamics Fluoroimager 595, excitation 488, emission 530. The resulting images are shown in Fig. 107.

In Fig. 107, Panel A shows the results of the target titration when no Arrestor™ oligonucleotide was included in the secondary reaction; Panel B shows the results of the same target titration using an Arrestor™ that extended 2 nt into the non-target complementary region of the primary probe; and Panel C shows the results of the same target titration using an Arrestor™ oligonucleotide that extended 4 nt into the non-target complementary region of the primary probe. The product of the secondary cleavage reaction is seen as a band near the bottom of each panel. The first two lanes of each panel (*i.e.*, 1 and 2, 11 and 12, 21 and 22) lacked target DNA, and the signal the co-migrates with the product band represents the nonspecific background under each set of conditions.

It can be seen by visual inspection of these panels that the background signal is both reduced, and made more predictable, by the inclusion of either species of Arrestor™ oligonucleotides. In addition to reducing the background in the no-target control lanes, the background reduction in the reactions that had the more dilute amounts of target included is reduced, leading to a signal that is a more accurate reflection of the target contained within the reaction, thus improving the quantitative range of the multiple, sequential invasive cleavage reaction.

To quantify the impact of including the Arrestor™ oligonucleotide in the secondary cleavage reaction under these conditions, the average product band signal from the reactions having the largest amount of target (*i.e.*, averages of the signals from lanes 9 and 10, lanes 19 and 20, and lanes 29 and 30), were compared to the averaged signal from the no-target contol lanes for each panel, determine the "fold over background," the factor of signal amplification over background, under each set of conditions. For the reactions without the Arrestor™, Panel A, the fold over background was 5.3; for Panel B, the fold over background

was 12.7; and for Panel C, the fold over background was 13.4, indicating that in this system inclusion of any Arrestor™ at least doubled the specificity of the signal over the Arrestor™-less reactions, and that the Arrestor™ that extended slightly farther into the non-target complementary region may be slightly more effective, at least in this embodiment of the system. This clearly shows the benefits of using an Arrestor™ to enhance the specificity of these reactions, an advantage that is of particular benefit at low levels of target nucleic acid.

EXAMPLE 50

"Arrestor™" Oligonucleotides Allow use of Higher Concentrations of Primary Probe Without Increasing Background Signal

It was demonstrated in Example 36, that increasing the concentration of the probe in the invasive cleavage reaction could dramatically increase the amount of signal generated for a given amount of target DNA. While not intending to limit the explanation to any specific mechanism, this is believed to be caused by the fact that increased concentration of probe increases the rate at which the cleaved probe is supplanted by an uncleaved copy, thereby increasing the apparent turnover rate of the cleavage reaction. Unfortunately, this effect could not heretofore be applied in the primary cleavage reaction of a multiple sequential Invader™ assay because the residual uncleaved primary probe can hybridize to the secondary target, in competition with the cleaved molecules, thereby reducing the efficacy of the secondary reaction. Elevated concentrations of primary probe exacerbate this problem. Further, the resulting complexes, as described above, can be cleaved at a low level, contributing to background. Therefore, increasing the primary probe can have the double negative effect of both slowing the secondary reaction and increasing the level of this form of non target-specific background. The use of an Arrestor™ to sequester or neutralize the residual primary probe allows this concentration-enhancing effect to be applied to these sequential reactions.

To demonstrate this effect, two sets of reactions were conducted. In the first set of reactions, the reactions were conducted using a range of primary probe concentrations, but no Arrestor™ oligonucleotide was supplied in the secondary reaction. In the second set of reactions, the same probe concentrations were used, but an Arrestor™ was added for the secondary reactions.

All reactions were performed in duplicate. Primary Invader™ reactions were done in a final volume of 10 μl and contained: 10 mM MOPS, pH 7.5, 7.5 mM $MgCl_2$, 500 fm of primary Invader™ (218-55-05; SEQ ID NO:171); 30 ng of AfuFEN1 enzyme and 10 ng of

human genomic DNA. 100 zeptomoles of HBV pAM6 #2 amplicon was included in all even numbered reactions (by reference to Figs. 108A and B). Reactions included 10 pmoles, 20 pmoles, 50 pmoles, 100 pmoles or 150 pmoles of primary probe (218-55-02; SEQ ID NO:170). MOPS, target and Invader™ oligonucleotides were combined to a final volume of 7 μl. Samples were heat denatured at 95°C for 5 minutes, then cooled to 67°C. During the 5 minute denaturation, $MgCl_2$, probe and enzyme were combined. The primary Invader™ reactions were initiated by the addition of 3 μl of $MgCl_2$, probe and enzyme mix, to the final concentrations indicated above. Reactions were incubated for 30 minutes at 67°C. The reactions were then cooled to 52°C, and each primary Invader™ reaction received the following secondary reaction components in a total volume of 4 μl: 2.5 pmoles secondary target (oligo number 218-95-04; SEQ ID NO:172); 10 pmoles secondary probe (oligo number 228-48-04; SEQ ID NO:173). The reactions that included the Arrestor™ oligonucleotide had either 40 pmoles, 80 pmoles, 200 pmoles, 400 pmoles or 600 pmoles of Arrestor™ (oligo number 218-95-01; SEQ ID NO:174), added at a 4-fold molar excess over the primary probe amount for each reaction, with this mix. Reactions were then incubated at 52°C for 30 minutes. The reactions were stopped by the addition of 10 μl of a solution of 95% formamide, 10 mM EDTA and 0.02% crystal violet. All samples were heated to 95°C for 1 minute, and 4 μl of each sample were resolved by electrophoresis through 20% denaturing acrylamide gel (19:1 cross-linked) with 7 M urea, in a buffer containing 45 mM Tris-Borate (pH8.3) and 1.4 mM EDTA. The results were imaged using the Molecular Dynamics Fluoroimager 595, excitation 488, emission 530. The resulting images for the reactions either without or with an Arrestor™ oligonucleotide are shown in Figs. 108A and 108B, respectively. The products of cleavage of the secondary probe are seen as a band near the bottom of each panel.

In Fig. 108A, lane sets 1 and 2 show results with 10 pmoles of primary probe; 3 and 4 had 20 pmoles; 5 and 6 had 50 pmoles; 7 and 8 had 100 pmoles; and 9 and 10 had 150 pmoles. It can be seen by visual examination, that the increases in the amount of primary probe have the combined effect of slightly increasing the background in the no-target lanes (odd numbers) while reducing the specific signal in the presence of target (even numbered lanes), and therefore the reducing the specificity of the reaction if viewed as the measure of "fold over background," demonstrating that the approach of increasing signal by increasing probe cannot be applied in these sequential reactions.

In Fig. 108B, lane sets 1 and 2 show results with 10 pmoles of primary probe; while 3 and 4 had 20 pmoles; 5 and 6 had 50 pmoles; 7 and 8 had 100 pmoles; and 9 and 10 had 150 pmoles. In addition, each reaction included 4-fold molar excess of the Arrestor™ oligonucleotide added before the secondary cleavage reaction. It can be seen by visual examination that the background in the no-target lanes (odd numbers) is lower in all cases, while the specific signal in the presence of target (even numbered lanes) increases with increased amounts of primary probe, leading to a greater "fold over background" sensitivity at this target level.

To quantitatively compare these effects, the fluorescence signal from the products of both non-specific and specific cleavage were measured. The results are depicted graphically in Fig. 108C, graphed as a measure of the percentage of the secondary probe cleaved during the reaction, compared to the amount of primary probe used. Examination of the plots from the no-target reactions confirms that the background in the absence of the Arrestor™ is, in general, roughly two-fold higher, and that both increase slightly with the increasing probe amounts. The specific signals however, diverge between the two sets of reaction more dramatically. While the signal in the no-Arrestor™ reactions decreases steadily as primary probe was increased, the signal in the Arrestor™ reactions continued to increase. At the highest primary probe concentrations tested, the no-Arrestor™ reactions had specific signal that was only 1.7 fold over background, while the Arrestor™ reactions detected the 100 zmoles (60,000 copies) of target with a signal 6.5 fold over background, thus demonstrating the improvement in the sequential invasive cleavage reaction when an Arrestor™ oligonucleotide is included.

EXAMPLE 51

Modified Backbones Improve Performance of Arrestor™ Oligonucleotides

All Natural "Arrestor™" Oligo With No 3'-Amine

The reactions described in the previous two Examples used Arrestor™ oligonucleotides that were constructed using 2' O-methyl ribose backbone, and which included a positively charged amine group on the 3' terminal nucleotide. The modifications were made specifically to reduce enzyme interaction with the primary probe/Arrestor™ complex. During the development of the present invention, it was determined that the 2' O-methy modified oligonucleotides are somewhat resistant to cleavage by the 5' nucleases, just

as they are slowly degraded by nucleases when used in antisense applications (*See e.g.*, Kawasaki *et al.*, J. Med. Chem., 36:831 [1993]).

Further, as demonstrated in Example 35, the presence of an amino group on the 3' end of an oligonucleotide reduces its ability to direct invasive cleavage. To reduce the possibility that the Arrestor™ oligonucleotide would form a cleavage structure in this way, an amino group was included in the design of the experiments described in this and other Examples.

Initial designs of the Arrestor™ oligonucleotides (sometimes referred to as "blockers") did not include these modifications, and these molecules were found to provide no benefit in reducing background cleavage in the sequential invasive cleavage assay and, in fact, sometimes contributed to background by inducing cleavage at an unanticipated site, presumably by providing some element to an alternative cleavage structure. The effects of natural and modified Arrestor™s on the background noise in these reactions are examined in this Example.

The efficacy of an "all-natural Arrestor™ (*i.e.*, an Arrestor™ that did not contain any base analogs or modifications) was examined by comparison to an identical reactions that lacked Arrestors™. All reactions were performed in duplicate, and were conducted as follows. Two master mixes were assembled, each containing 12.5 mM MOPS, pH 7.5, 500 fmoles primary Invader™ oligonucleotide #218-55-05 (SEQ ID NO:171), 10 ng human genomic DNA (Novagen) and 30 ng AfuFEN1 enzyme for every 8 μl of mix. Mix A contained no added HBV genomic amplicon DNA, mix B contained 600,000 molecules of HBV genomic amplicon DNA, pAM6 #2. The mixes were distributed to the reaction tubes, in aliquots of 8 μl/tube as follows: mix A to tubes 1, 2, 5 and 6; and mix B to tubes 3, 4, 7 and 8. The samples were incubated at 95°C for 4 minutes to denature the HBV genomic amplicon DNA. The reactions were then cooled to 67°C and 2ul of a mix containing 37.5 mM $MgCl_2$ and 10 pmoles 218-55-02B (SEQ ID NO:185) for every 2 μl, was added to each sample. The samples were then incubated at 67°C for 30 minutes. Two secondary reaction master mixes were prepared, each containing 10 pmoles of secondary probe oligo #228-48-04N (SEQ ID NO:178) and 2.5 pmoles of secondary target oligonucleotide #218-95-04 (SEQ ID NO:172) for every 3 μl of mix. Mix 2A contained no additional oligonucleotide, while mix 2B contained 50 pmoles of the natural "Arrestor™" oligonucleotide #241-62-02 (SEQ ID NO:186). After the initial 30 minute incubation at 67°C, the temperature was adjusted to 52°C, and 3 μl of a secondary reaction mix was added to each sample, as follows: Mix 2A was added to samples #1-4; and Mix 2B was added to samples #5-8. The samples were then

incubated for 30 minutes at 52°C. The reactions were then stopped by the addition of 10 μl of a solution of 95% formamide, 10 mM EDTA and 0.02% crystal violet.

All of the samples were heated to 95°C for 2 minutes, and 4 μl of each sample were resolved by electrophoresis through a 20% denaturing acrylamide gel (19:1 cross-linked) with 7 M urea, in a buffer containing 45 mM Tris-Borate (pH8.3) and 1.4 mM EDTA. The results were imaged using the Molecular Dynamics Fluoroimager 595, excitation 488, emission 530. The resulting image is shown in Fig. 109A.

To compare the effects of the various modifications made to the Arrestors™, reactions were performed using Arrestors™ having all natural bases, but including a 3' terminal amine; Arrestors™ having the 3' portion composed of 2' O-methyl nucleotides, plus the 3' terminal amine; and Arrestors™ composed entirely of 2' O-methyl nucleotides, plus the 3' terminal amine. These were compared to reactions performed without an Arrestor™. The reactions were conducted as follows. Two master mixes were assembled, all mixes contained 14.3 mM MOPS, pH 7.5, 500 fmoles primary Invader™ oligo #218-55-05 (SEQ ID NO:171) and 10 ng human genomic DNA (Novagen) for every 7 μl of mix. Mix A contained no added HBV genomic amplicon DNA, mix B contained 600,000 molecules of HBV genomic amplicon DNA, pAM6 #2. The mixes were distributed to the reaction tubes, at 7 μl/tube: mix A to tubes 1, 2, 5, 6, 9, 10, 13 and 14; and mix B to tubes 3, 4, 7, 8, 11, 12, 15 and 16. The samples were warmed to 95°C for 4 minutes to denature the HBV DNA. The reactions were then cooled to 67°C and 3 μl of a mix containing 25 mM $MgCl_2$, 25 pmoles 218-55-02B (SEQ ID NO:185) and 30 ng AfuFEN1 enzyme per 3 μl, were added to each sample. The samples were then incubated at 67°C for 30 minutes. Four secondary reaction master mixes were prepared; all mixes contained 10 pmoles of secondary probe oligonucleotide #228-48-04B (SEQ ID NO:190) and 2.5 pmoles of secondary target oligonucleotide #218-95-04 (SEQ ID NO:172) for every 3 μl of mix. Mix 2A contained no additional oligonucleotide, while mix 2B contained 100 pmoles of the natural+amine Arrestor™ oligonucleotide # 241-62-01 (SEQ ID NO:187), mix 2C contained 100 pmoles of partially O-methyl+amine oligonucleotide # 241-62-03 (SEQ ID NO:188) and mix 2D contained 100 pmoles of all O-methyl+amine oligonucleotide # 241-64-01 (SEQ ID NO:189). After the initial 30 minute incubation at 67°C, the temperature was adjusted to 52°C and 3 μl of a secondary reaction mix was added to each sample, as follows: mix 2A was added to samples #1-4; mix 2B was added to samples #5-8; mix 2C was added to samples #9-12; and mix 2D was added to

samples #13-16. The samples were incubated for 30 minutes at 52°C, then stopped by the addition of 10 μl of a solution of 95% formamide, 10 mM NaEDTA, and 0.2% crystal violet.

All samples were heated to 95°C for 2 minutes, and 4 μl of each sample were resolved by electrophoresis through a 20% denaturing acrylamide gel (19:1 cross-linked) with 7 M urea, in a buffer containing 45 mM Tris-Borate (pH8.3) and 1.4 mM EDTA. The results were imaged using the Molecular Dynamics Fluoroimager 595, excitation 488, emission 530. The resulting image is shown in Fig. 109B.

In Fig. 109A, the left hand panel shows the reactions that lacked an Arrestor™, while the right hand panel shows the data from reactions that included the all natural Arrestor™ oligonucleotide. The first two lanes of each panel are from no-target controls, the second set of lanes contained target. The products of cleavage are visible in the bottom one/fourth of each panel. The position at which the specific reaction products should run is indicated by arrows on left and right.

It can be seen by examination of these data, that the reactions run in the absence of Arrestor™ show reproducible quality between the replicates, and show significant cleavage only when target is present. In contrast, the addition of another unmodified oligonucleotide into the reactions causes great variation between the replicate lanes (*e.g.*, lanes 5 and 6 were provided with the same reactants, but produced markedly different results). The introduction of the all natural Arrestor™ produced, rather than reduced, background in these no-target lanes, and increased cleavage at other sites (*i.e.*, the bands other that those indicated by the arrows flanking the panels). For these reasons the modifications which are described above, the effects of which are shown on Fig. 109B, were incorporated.

The first 4 lanes of Fig. 109B show the products of duplicate reactions without an Arrestor™, plus or minus the HBV target (lanes 1, 2, and lanes 3, 4, respectively); The next 4 lanes, 5, 6 and 7, 8 used a natural Arrestor™ oligonucleotide having a 3' terminal amine; lanes 9, 10 and 11, 12 used the Arrestor™ with a 3' portion composed of 2' O-methyl nucleotides, and having a 3' terminal amine; lanes 13, 14 and 15, 16 used the Arrestor™ composed entirely of 2' O-methyl nucleotides and having a 3' terminal amine. The products of cleavage of the secondary probe are visible in the lower one third of each panel.

Visual inspection of these data shows that the addition of the 3' terminal amine to the natural Arrestor™ suppresses the aberrant cleavage seen in Fig. 109A, but this Arrestor™ does not improve the performance of the reaction, as compared to the no-Arrestor™ controls. In contrast, the use of the 2' O-methyl nucleotides in the body of the Arrestor™

oligonucleotide does reduce background, whether partially or completely substituted. To quantify the relative effects of these modifications, the fluorescence from each of the co-migrating product bands was measured, the signals from the duplicate lanes were averaged and the "fold over background" was calculated for each reaction containing target nucleic acid.

When Arrestor™ was omitted, the target specific signal (lanes 3, 4) was 27-fold over the no target background; the natural Arrestor™+amine gave a signal of 17-fold over background; the partial 2' O-methyl+ amine gave a signal of 47-fold over background; and the completely 2' O-methyl+ amine gave a signal of 33 fold over background.

These Figures show that both modifications can have a beneficial effect on the specificity of the multiple, sequential invasive cleavage assay. They also show that the use of the 2' O-methyl substituted backbone, either partial or entire, markedly improves the specificity of these reactions. It is intended that in various embodiments of the present inventon, that any number of modifications that make either the Arrestor™ or the complex it forms with the primary target resistant to nucleases will provide similar enhancement.

EXAMPLE 52

Effect of Arrestor™ Length on Signal Enhancement in Multiple Sequential Invasive Cleavage Assays

As noted in the Description of the Invention, the optimal length for an Arrestor™ oligonucleotide depends upon the design of the other nucleic acid elements of the Invader™ reaction, particularly on the design of the primary probe. In this Example, the effects of varying the length of the Arrestor™ oligonucleotide were explored in systems using two different secondary probes. A schematic diagram showing these Arrestors™ aligned as they would hybridize to the primary probe oligonucleotide is provided in Fig. 110C. In this Figure, the region of the primary probe that recognizes the target nucleic acid is shown underlined; the non-underlined portion, plus the first underlined base is the portion that is released by the first cleavage, and goes on to participate in the second or subsequent cleavage structure.

All reactions were performed in duplicate. The Invader™ reactions were done in a final volume of 10 µl final volume containing 10 mM MOPS, pH 7.5, mM $MgCl_2$, 500 fmoles of primary Invader™ 241-95-01, (SEQ ID NO:176), 25 pmoles of primary probe 241-95-02 (SEQ ID NO:175), 30 ng of AfuFEN1 enzyme, and 10 ng of human genomic DNA,

and if included, 1 amoles of HBV amplicon pAM 6 #2. MOPS, target DNA, and Invader™ oligonucleotides were combined to a final volume of 7 µl. Samples were heat denatured at 95°C for 5 minutes, then cooled to 67°C. During the 5 minute denaturation, $MgCl_2$, probe and enzyme were combined. The primary Invader™ reactions were initiated by the addition of 3 µl of $MgCl_2$, probe and enzyme mix, to the final concentrations indicated above. Reactions were incubated for 30 minutes at 67°C. The reaction were then cooled to 52°C, and each primary Invader™ reaction received the following secondary reaction components in a total volume of 3 µl: 2.5 pmoles secondary target 241-95-07 (SEQ ID NO:177), 10 pmoles of either secondary probe 228-48-04 (SEQ ID NO:173), or 228-48-04N (SEQ ID NO:178) and 100 pmoles of an Arrestor™ oligonucleotide, either 241-95-03 (SEQ ID NO:179), 241-95-04 (SEQ ID NO:180), 241-95-05 (SEQ ID NO:181) or 241-95-06 (SEQ ID NO:182). The Arrestors™ were omitted from some reactions as controls for Arrestor™ effects.

The reactions were incubated at 52°C for 34 minutes, and were then stopped by the addition of 10 µl of 95% formamide, 10 mM EDTA, and 0.02% crystal violet. All samples were heated to 95°C for 1 minute, and 4 µl of each sample were resolved by electrophoresis through 20% denaturing acrylamide gel (19:1 cross-linked) with 7 M urea, in a buffer containing 45 mM Tris-Borate (pH8.3) and 1.4 mM EDTA. The results were imaged using the Molecular Dynamics Fluoroimager 595, excitation 488, emission 530. The resulting images for the reactions with the shorter and longer secondary probes are shown in Figs. 110A and 110B, respectively.

In each Figure, the products of cleavage are visible as bands in the bottom half of each lane. The first 4 lanes of each Figure show the products of duplicate reactions without an Arrestor™, plus or minus the HBV target (lanes sets 1 and 2 respectively); in the next 4 lanes, sets 3 and 4 used the shortest Arrestor™ 241-95-03 (SEQ ID NO:179); lanes 5 and 6 used 241-95-04 (SEQ ID NO:180); lanes 7 and 8 used 241-95-05 (SEQ ID NO:181); and lanes 9 and 10 used 241-95-06 (SEQ ID NO:182).

The principal background of concern is the band that appears in the "no target" control lanes (odd numbers; this band co-migrates with the target-specific signal near the bottom of each gel panel). Visual inspection shows that the shortest Arrestor™ was the least effective at suppressing this background, and that the efficacy was increased when the Arrestor™ extended further into the portion that participates in the subsequent cleavage reaction. Even with this difference in effect, it can be seen from these data that there is much latitude in the design of the Arrestor™ oligonucleotide. The choice of lengths will be influenced by the

temperature at which the reaction making use of the Arrestor™ is performed, the lengths of the duplexes formed between the primary probe and the target, the primary probe and the secondary target, and the relative concentrations of the different nucleic acid species in the reactions.

EXAMPLE 53

Effect of Arrestor™ Concentration on Signal Enhancement in Multiple Sequential Invasive Cleavage Assays

In examining the effects of including Arrestor™ oligonucleotides in these cleavage reactions, it was of interest to determine if the concentration of the Arrestor™ in excess of the primary probe concentration would have an effect on yields of either non-specific or specific signal, and if the length of the Arrestor™ would be a factor. These two variable were investigated in the following Example.

All reactions were performed in duplicate. The primary Invader™ reactions were done in a final volume of 10 µl and contained 10 mM MOPS, pH 7.5; 7.5 mM $MgCl_2$, 500 fmoles of primary Invader™ 241-95-01 (SEQ ID NO:176), 25 pmoles of primary probe 241-95-02 (SEQ ID NO:175), 30 ng of AfuFEN1 enzyme, and 10 ng of human genomic DNA. Where included, the target DNA was 1 amole of HBV amplicon pAM 6 #2, as described above. MOPS, target and Invader™ were combined to a final volume of 7 µl. The samples were heat denatured at 95°C for 5 minutes, then cooled to 67°C. During the 5 minute denaturation, $MgCl_2$, probe and enzyme were combined. The primary Invader™ reactions were initiated by the addition of 3 µl of $MgCl_2$, probe and enzyme mix. The reactions were incubated for 30 minutes at 67°C. The reactions were then cooled to 52°C and each primary Invader™ reaction received the following secondary reaction components: 2.5 pmoles secondary target 241-95-07 (SEQ ID NO:177), 10 pmoles secondary probe 228-48-04 (SEQ ID NO:173); and, if included, 50, 100 or 200 pmoles of either Arrestor™ 241-95-03 (SEQ ID NO:179) or 241-95-05 (SEQ ID NO:181), in a total volume of 3 µl. Reactions were then incubated at 52°C for 35 minutes. Reactions were stopped by the addition of 10 µl of 95% formamide, 10 mM EDTA, and 0.02% crystal violet. All of the samples were heated to 95°C for 1 minute, and 4 µl of each sample were resolved by electrophoresis through 20% denaturing acrylamide gel (19:1 cross-linked) with 7 M urea, in a buffer containing 45 mM Tris-Borate (pH8.3), and 1.4 mM EDTA. The results were imaged using the Molecular

Dynamics Fluoroimager 595, excitation 488, emission 530. The resulting images are shown as a composite image in Fig. 111.

Each of the duplicate reactions were loaded on the gel in adjacent lanes and are labeled with a single lane number. All odd numbered lanes were no-target controls. Lanes 1 and 2 had no Arrestor™ oligonucleotide added; lanes 3-8 show results from reactions containing the shorter Arrestor™, 241-95-03 (SEQ ID NO:179); lanes 9-14 show results from reactions containing the longer Arrestor™, 241-95-05 (SEQ ID NO:181). The products of cleavage from the secondary reaction are visible in the bottom one third of each panel. Visual inspection of these data (*i.e.*, comparison of the specific products to the background bands) shows that both Arrestors™ have some beneficial effect at all concentration.

To quantify the relative effects of Arrestor™ length and concentration, the fluorescence from each of the co-migrating product bands was measured, the signals from the duplicate lanes were averaged and the "fold over background" (signal+target/signal-target) was calculated for each reaction containing target nucleic acid. The reaction lacking an Arrestor™ yielded a signal approximately 27-fold over background. Inclusion of the shorter Arrestor™ at 50, 100 or 200 pmoles produced products at 42, 51 and 60-fold over background, respectively. This shows that while the short arrestr at the lowest concentration seems to be less effective than the longer Arrestors™ (*See*, previous Example) this can be compensated for by increasing the concentration of Arrestor™, and thereby the Arrestor™:primary probe ratio.

In contrast, inclusion of the longer Arrestor™ at 50, 100 or 200 pmoles produced products at 60, 32 and 24 fold over background, respectively. At the lowest concentration, the efficacy of this longer Arrestor™ relative to the shorter Arrestor™ is consistent with the previous Example. Increasing the concentration, however, decreased the yield of specific product, suggesting a competition effect with some element of the secondary cleavage reaction.

These data show that the Arrestor™ oligonucleotides can be used to advantage in a number of specific reaction designs. The choice of concentration will be influenced by the temperature at which the reaction making use of the Arrestor™ is performed, the lengths of the duplexes formed between the primary probe and the target, the primary probe and the secondary target, and between the primary probe and the Arrestor™.

Selection of oligonucleotides for target nucleic acids other than the HBV shown here, (*e.g.*, oligonucleotide composition and length), and the optimization of cleavage reaction

conditions in accord with the models provided here follow routine methods and common practice well known to those skilled in the methods of molecular biology.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Hall, Jeff G.
Lyamichev, Victor I.
Mast, Andrea L.
Brow, Mary Ann D.

(ii) TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple Sequential Invasive Cleavages

(iii) NUMBER OF SEQUENCES: 190

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Medlen & Carroll, LLP
(B) STREET: 220 Montgomery Street, Suite 2200
(C) CITY: San Francisco
(D) STATE: California
(E) COUNTRY: United States Of America
(F) ZIP: 94104

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US97/01072
(B) FILING DATE: 21-JAN-1997

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/759,038
(B) FILING DATE: 02-DEC-1996

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/758,314
(B) FILING DATE: 02-DEC-1996

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/756,386
(B) FILING DATE: 29-NOV-1996

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/682,853
(B) FILING DATE: 12-JUL-1996

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/599,491
(B) FILING DATE: 24-JAN-1996

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/823,516
(B) FILING DATE: 24-MAR-1997

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: MacKnight, Kamrin T.
(B) REGISTRATION NUMBER: 38,230
(C) REFERENCE/DOCKET NUMBER: FORS-03295

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (415) 705-8410
(B) TELEFAX: (415) 397-8338

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2506 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
ATGAGGGGGA TGCTGCCCCT CTTTGAGCCC AAGGGCCGGG TCCTCCTGGT GGACGGCCAC     60
CACCTGGCCT ACCGCACCTT CCACGCCCTG AAGGGCCTCA CCACCAGCCG GGGGGAGCCG    120
GTGCAGGCGG TCTACGGCTT CGCCAAGAGC CTCCTCAAGG CCCTCAAGGA GGACGGGGAC    180
GCGGTGATCG TGGTCTTTGA CGCCAAGGCC CCCTCCTTCC GCCACGAGGC CTACGGGGGG    240
TACAAGGCGG GCCGGGCCCC CACGCCGGAG GACTTTCCCC GGCAACTCGC CCTCATCAAG    300
GAGCTGGTGG ACCTCCTGGG GCTGGCGCGC CTCGAGGTCC CGGGCTACGA GGCGGACGAC    360
GTCCTGGCCA GCCTGGCCAA GAAGGCGGAA AAGGAGGGCT ACGAGGTCCG CATCCTCACC    420
GCCGACAAAG ACCTTTACCA GCTCCTTTCC GACCGCATCC ACGTCCTCCA CCCCGAGGGG    480
TACCTCATCA CCCCGGCCTG GCTTTGGGAA AAGTACGGCC TGAGGCCCGA CCAGTGGGCC    540
GACTACCGGG CCCTGACCGG GGACGAGTCC GACAACCTTC CCGGGGTCAA GGGCATCGGG    600
GAGAAGACGG CGAGGAAGCT TCTGGAGGAG TGGGGGAGCC TGGAAGCCCT CCTCAAGAAC    660
CTGGACCGGC TGAAGCCCGC CATCCGGGAG AAGATCCTGG CCCACATGGA CGATCTGAAG    720
CTCTCCTGGG ACCTGGCCAA GGTGCGCACC GACCTGCCCC TGGAGGTGGA CTTCGCCAAA    780
AGGCGGGAGC CCGACCGGGA GAGGCTTAGG GCCTTTCTGG AGAGGCTTGA GTTTGGCAGC    840
CTCCTCCACG AGTTCGGCCT TCTGGAAAGC CCCAAGGCCC TGGAGGAGGC CCCCTGGCCC    900
CCGCCGGAAG GGGCCTTCGT GGGCTTTGTG CTTTCCCGCA AGGAGCCCAT GTGGGCCGAT    960
CTTCTGGCCC TGGCCGCCGC CAGGGGGGGC CGGGTCCACC GGGCCCCCGA GCCTTATAAA   1020
GCCCTCAGGG ACCTGAAGGA GGCGCGGGGG CTTCTCGCCA AAGACCTGAG CGTTCTGGCC   1080
CTGAGGGAAG GCCTTGGCCT CCCGCCCGGC GACGACCCCA TGCTCCTCGC CTACCTCCTG   1140
GACCCTTCCA ACACCACCCC CGAGGGGGTG GCCCGGCGCT ACGGCGGGGA GTGGACGGAG   1200
GAGGCGGGGG AGCGGGCCGC CCTTTCCGAG AGGCTCTTCG CCAACCTGTG GGGGAGGCTT   1260
GAGGGGGAGG AGAGGCTCCT TTGGCTTTAC CGGGAGGTGG AGAGGCCCCT TTCCGCTGTC   1320
CTGGCCCACA TGGAGGCCAC GGGGGTGCGC CTGGACGTGG CCTATCTCAG GGCCTTGTCC   1380
```

```
CTGGAGGTGG CCGAGGAGAT CGCCCGCCTC GAGGCCGAGG TCTTCCGCCT GGCCGGCCAC      1440
CCCTTCAACC TCAACTCCCG GGACCAGCTG GAAAGGGTCC TCTTTGACGA GCTAGGGCTT      1500
CCCGCCATCG GCAAGACGGA GAAGACCGGC AAGCGCTCCA CCAGCGCCGC CGTCCTGGAG      1560
GCCCTCCGCG AGGCCCACCC CATCGTGGAG AAGATCCTGC AGTACCGGGA GCTCACCAAG      1620
CTGAAGAGCA CCTACATTGA CCCCTTGCCG GACCTCATCC ACCCCAGGAC GGGCCGCCTC      1680
CACACCCGCT TCAACCAGAC GGCCACGGCC ACGGGCAGGC TAAGTAGCTC CGATCCCAAC      1740
CTCCAGAACA TCCCCGTCCG CACCCCGCTT GGGCAGAGGA TCCGCCGGGC CTTCATCGCC      1800
GAGGAGGGGT GGCTATTGGT GGCCCTGGAC TATAGCCAGA TAGAGCTCAG GGTGCTGGCC      1860
CACCTCTCCG GCGACGAGAA CCTGATCCGG GTCTTCCAGG AGGGGCGGGA CATCCACACG      1920
GAGACCGCCA GCTGGATGTT CGGCGTCCCC CGGGAGGCCG TGGACCCCCT GATGCGCCGG      1980
GCGGCCAAGA CCATCAACTT CGGGGTCCTC TACGGCATGT CGGCCCACCG CCTCTCCCAG      2040
GAGCTAGCCA TCCCTTACGA GGAGGCCCAG GCCTTCATTG AGCGCTACTT TCAGAGCTTC      2100
CCCAAGGTGC GGGCCTGGAT TGAGAAGACC CTGGAGGAGG GCAGGAGGCG GGGGTACGTG      2160
GAGACCCTCT TCGGCCGCCG CCGCTACGTG CCAGACCTAG AGGCCCGGGT GAAGAGCGTG      2220
CGGGAGGCGG CCGAGCGCAT GGCCTTCAAC ATGCCCGTCC AGGGCACCGC CGCCGACCTC      2280
ATGAAGCTGG CTATGGTGAA GCTCTTCCCC AGGCTGGAGG AAATGGGGGC CAGGATGCTC      2340
CTTCAGGTCC ACGACGAGCT GGTCCTCGAG GCCCCAAAAG AGAGGGCGGA GGCCGTGGCC      2400
CGGCTGGCCA AGGAGGTCAT GGAGGGGGTG TATCCCCTGG CCGTGCCCCT GGAGGTGGAG      2460
GTGGGGATAG GGGAGGACTG GCTCTCCGCC AAGGAGTGAT ACCACC                     2506
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2496 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
ATGGCGATGC TTCCCCTCTT TGAGCCCAAA GGCCGCGTGC TCCTGGTGGA CGGCCACCAC        60
CTGGCCTACC GCACCTTCTT TGCCCTCAAG GGCCTCACCA CCAGCCGCGG CGAACCCGTT       120
CAGGCGGTCT ACGGCTTCGC CAAAAGCCTC CTCAAGGCCC TGAAGGAGGA CGGGGACGTG       180
GTGGTGGTGG TCTTTGACGC CAAGGCCCCC TCCTTCCGCC ACGAGGCCTA CGAGGCCTAC       240
AAGGCGGGCC GGGCCCCCAC CCCGGAGGAC TTTCCCCGGC AGCTGGCCCT CATCAAGGAG       300
TTGGTGGACC TCCTAGGCCT TGTGCGGCTG GAGGTTCCCG GCTTTGAGGC GGACGACGTG       360
CTGGCCACCC TGGCCAAGCG GGCGGAAAAG GAGGGGTACG AGGTGCGCAT CCTCACTGCC       420
GACCGCGACC TCTACCAGCT CCTTTCGGAG CGCATCGCCA TCCTCCACCC TGAGGGGTAC       480
CTGATCACCC CGGCGTGGCT TTACGAGAAG TACGGCCTGC GCCCGGAGCA GTGGGTGGAC       540
```

TACCGGGCCC TGGCGGGGGA CCCCTCGGAT AACATCCCCG GGGTGAAGGG CATCGGGGAG 600
AAGACCGCCC AGAGGCTCAT CCGCGAGTGG GGGAGCCTGG AAAACCTCTT CCAGCACCTG 660
GACCAGGTGA AGCCCTCCTT GCGGGAGAAG CTCCAGGCGG GCATGGAGGC CCTGGCCCTT 720
TCCCGGAAGC TTTCCCAGGT GCACACTGAC CTGCCCCTGG AGGTGGACTT CGGGAGGCGC 780
CGCACACCCA ACCTGGAGGG TCTGCGGGCT TTTTTGGAGC GGTTGGAGTT TGGAAGCCTC 840
CTCCACGAGT TCGGCCTCCT GGAGGGGCCG AAGGCGGCAG AGGAGGCCCC CTGGCCCCCT 900
CCGGAAGGGG CTTTTTTGGG CTTTTCCTTT TCCCGTCCCG AGCCCATGTG GGCCGAGCTT 960
CTGGCCCTGG CTGGGGCGTG GGAGGGGCGC CTCCATCGGG CACAAGACCC CCTTAGGGGC 1020
CTGAGGGACC TTAAGGGGGT GCGGGGAATC CTGGCCAAGG ACCTGGCGGT TTTGGCCCTG 1080
CGGGAGGGCC TGGACCTCTT CCCAGAGGAC GACCCCATGC TCCTGGCCTA CCTTCTGGAC 1140
CCCTCCAACA CCACCCCTGA GGGGGTGGCC CGGCGTTACG GGGGGGAGTG GACGGAGGAT 1200
GCGGGGGAGA GGGCCCTCCT GGCCGAGCGC CTCTTCCAGA CCCTAAAGGA GCGCCTTAAG 1260
GGAGAAGAAC GCCTGCTTTG GCTTTACGAG GAGGTGGAGA AGCCGCTTTC CCGGGTGTTG 1320
GCCCGGATGG AGGCCACGGG GGTCCGGCTG GACGTGGCCT ACCTCCAGGC CCTCTCCCTG 1380
GAGGTGGAGG CGGAGGTGCG CCAGCTGGAG GAGGAGGTCT TCCGCCTGGC CGGCCACCCC 1440
TTCAACCTCA ACTCCCGCGA CCAGCTGGAG CGGGTGCTCT TTGACGAGCT GGGCCTGCCT 1500
GCCATCGGCA AGACGGAGAA GACGGGGAAA CGCTCCACCA GCGCTGCCGT GCTGGAGGCC 1560
CTGCGAGAGG CCCACCCCAT CGTGGACCGC ATCCTGCAGT ACCGGGAGCT CACCAAGCTC 1620
AAGAACACCT ACATAGACCC CCTGCCCGCC CTGGTCCACC CCAAGACCGG CCGGCTCCAC 1680
ACCCGCTTCA ACCAGACGGC CACCGCCACG GGCAGGCTTT CCAGCTCCGA CCCCAACCTG 1740
CAGAACATCC CCGTGCGCAC CCCTCTGGGC CAGCGCATCC GCCGAGCCTT CGTGGCCGAG 1800
GAGGGCTGGG TGCTGGTGGT CTTGGACTAC AGCCAGATTG AGCTTCGGGT CCTGGCCCAC 1860
CTCTCCGGGG ACGAGAACCT GATCCGGGTC TTTCAGGAGG GGAGGGACAT CCACACCCAG 1920
ACCGCCAGCT GGATGTTCGG CGTTTCCCCC GAAGGGGTAG ACCCTCTGAT GCGCCGGGCG 1980
GCCAAGACCA TCAACTTCGG GGTGCTCTAC GGCATGTCCG CCCACCGCCT CTCCGGGGAG 2040
CTTTCCATCC CCTACGAGGA GGCGGTGGCC TTCATTGAGC GCTACTTCCA GAGCTACCCC 2100
AAGGTGCGGG CCTGGATTGA GGGGACCCTC GAGGAGGGCC GCCGGCGGGG GTATGTGGAG 2160
ACCCTCTTCG GCCGCCGGCG CTATGTGCCC GACCTCAACG CCCGGGTGAA GAGCGTGCGC 2220
GAGGCGGCGG AGCGCATGGC CTTCAACATG CCGGTCCAGG GCACCGCCGC CGACCTCATG 2280
AAGCTGGCCA TGGTGCGGCT TTTCCCCCGG CTTCAGGAAC TGGGGGCGAG GATGCTTTTG 2340
CAGGTGCACG ACGAGCTGGT CCTCGAGGCC CCCAAGGACC GGGCGGAGAG GGTAGCCGCT 2400
TTGGCCAAGG AGGTCATGGA GGGGGTCTGG CCCCTGCAGG TGCCCCTGGA GGTGGAGGTG 2460
GGCCTGGGGG AGGACTGGCT CTCCGCCAAG GAGTAG 2496

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2504 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGGAGGCGA TGCTTCCGCT CTTTGAACCC AAAGGCCGGG TCCTCCTGGT GGACGGCCAC 60
CACCTGGCCT ACCGCACCTT CTTCGCCCTG AAGGGCCTCA CCACGAGCCG GGGCGAACCG 120
GTGCAGGCGG TCTACGGCTT CGCCAAGAGC CTCCTCAAGG CCCTGAAGGA GGACGGGTAC 180
AAGGCCGTCT TCGTGGTCTT TGACGCCAAG GCCCCCTCCT TCCGCCACGA GGCCTACGAG 240
GCCTACAAGG CGGGGAGGGC CCCGACCCCC GAGGACTTCC CCCGGCAGCT CGCCCTCATC 300
AAGGAGCTGG TGGACCTCCT GGGGTTTACC CGCCTCGAGG TCCCCGGCTA CGAGGCGGAC 360
GACGTTCTCG CCACCCTGGC CAAGAAGGCG GAAAAGGAGG GGTACGAGGT GCGCATCCTC 420
ACCGCCGACC GCGACCTCTA CCAACTCGTC TCCGACCGCG TCGCCGTCCT CCACCCCGAG 480
GGCCACCTCA TCACCCCGGA GTGGCTTTGG GAGAAGTACG GCCTCAGGCC GGAGCAGTGG 540
GTGGACTTCC GCGCCCTCGT GGGGGACCCC TCCGACAACC TCCCCGGGGT CAAGGGCATC 600
GGGGAGAAGA CCGCCCTCAA GCTCCTCAAG GAGTGGGGAA GCCTGGAAAA CCTCCTCAAG 660
AACCTGGACC GGGTAAAGCC AGAAAACGTC CGGGAGAAGA TCAAGGCCCA CCTGGAAGAC 720
CTCAGGCTCT CCTTGGAGCT CTCCCGGGTG CGCACCGACC TCCCCCTGGA GGTGGACCTC 780
GCCCAGGGGC GGGAGCCCGA CCGGGAGGGG CTTAGGGCCT TCCTGGAGAG GCTGGAGTTC 840
GGCAGCCTCC TCCACGAGTT CGGCCTCCTG GAGGCCCCCG CCCCCCTGGA GGAGGCCCCC 900
TGGCCCCCGC CGGAAGGGGC CTTCGTGGGC TTCGTCCTCT CCCGCCCCGA GCCCATGTGG 960
GCGGAGCTTA AAGCCCTGGC CGCCTGCAGG GACGGCCGGG TGCACCGGGC AGCAGACCCC 1020
TTGGCGGGGC TAAAGGACCT CAAGGAGGTC CGGGGCCTCC TCGCCAAGGA CCTCGCCGTC 1080
TTGGCCTCGA GGGAGGGGCT AGACCTCGTG CCCGGGGACG ACCCCATGCT CCTCGCCTAC 1140
CTCCTGGACC CCTCCAACAC CACCCCCGAG GGGGTGGCGC GGCGCTACGG GGGGGAGTGG 1200
ACGGAGGACG CCGCCCACCG GGCCCTCCTC TCGGAGAGGC TCCATCGGAA CCTCCTTAAG 1260
CGCCTCGAGG GGGAGGAGAA GCTCCTTTGG CTCTACCACG AGGTGGAAAA GCCCCTCTCC 1320
CGGGTCCTGG CCCACATGGA GGCCACCGGG GTACGGCTGG ACGTGGCCTA CCTTCAGGCC 1380
CTTTCCCTGG AGCTTGCGGA GGAGATCCGC CGCCTCGAGG AGGAGGTCTT CCGCTTGGCG 1440
GGCCACCCCT TCAACCTCAA CTCCCGGGAC CAGCTGGAAA GGGTGCTCTT TGACGAGCTT 1500
AGGCTTCCCG CCTTGGGGAA GACGCAAAAG ACAGGCAAGC GCTCCACCAG CGCCGCGGTG 1560
CTGGAGGCCC TACGGGAGGC CCACCCCATC GTGGAGAAGA TCCTCCAGCA CCGGGAGCTC 1620
ACCAAGCTCA AGAACACCTA CGTGGACCCC CTCCCAAGCC TCGTCCACCC GAGGACGGGC 1680
CGCCTCCACA CCCGCTTCAA CCAGACGGCC ACGGCCACGG GGAGGCTTAG TAGCTCCGAC 1740

```
CCCAACCTGC AGAACATCCC CGTCCGCACC CCCTTGGGCC AGAGGATCCG CCGGGCCTTC   1800
GTGGCCGAGG CGGGTTGGGC GTTGGTGGCC CTGGACTATA GCCAGATAGA GCTCCGCGTC   1860
CTCGCCCACC TCTCCGGGGA CGAAAACCTG ATCAGGGTCT TCCAGGAGGG GAAGGACATC   1920
CACACCCAGA CCGCAAGCTG GATGTTCGGC GTCCCCCCGG AGGCCGTGGA CCCCCTGATG   1980
CGCCGGGCGG CCAAGACGGT GAACTTCGGC GTCCTCTACG GCATGTCCGC CCATAGGCTC   2040
TCCCAGGAGC TTGCCATCCC CTACGAGGAG GCGGTGGCCT TTATAGAGGC TACTTCCAAA   2100
GCTTCCCCAA GGTGCGGGCC TGGATAGAAA AGACCCTGGA GGAGGGGAGG AAGCGGGGCT   2160
ACGTGGAAAC CCTCTTCGGA AGAAGGCGCT ACGTGCCCGA CCTCAACGCC CGGGTGAAGA   2220
GCGTCAGGGA GGCCGCGGAG CGCATGGCCT TCAACATGCC CGTCCAGGGC ACCGCCGCCG   2280
ACCTCATGAA GCTCGCCATG GTGAAGCTCT TCCCCCGCCT CCGGGAGATG GGGGCCCGCA   2340
TGCTCCTCCA GGTCCACGAC GAGCTCCTCC TGGAGGCCCC CCAAGCGCGG GCCGAGGAGG   2400
TGGCGGCTTT GGCCAAGGAG GCCATGGAGA AGGCCTATCC CCTCGCCGTG CCCCTGGAGG   2460
TGGAGGTGGG GATGGGGGAG GACTGGCTTT CCGCCAAGGG TTAG                    2504
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 832 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Arg Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu
1               5                   10                  15

Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys Gly
            20                  25                  30

Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala
        35                  40                  45

Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile Val
    50                  55                  60

Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly Gly
65                  70                  75                  80

Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu
                85                  90                  95

Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu Glu
            100                 105                 110

Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys Lys
        115                 120                 125

Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys Asp
    130                 135                 140

Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu Gly
145                 150                 155                 160
```

```
Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg Pro
                165                 170                 175

Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp Asn
            180                 185                 190

Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu Leu
        195                 200                 205

Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg Leu
    210                 215                 220

Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu Lys
225                 230                 235                 240

Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu Val
                245                 250                 255

Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala Phe
            260                 265                 270

Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu Leu
        275                 280                 285

Glu Ser Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu Gly
    290                 295                 300

Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala Asp
305                 310                 315                 320

Leu Leu Ala Leu Ala Ala Ala Arg Gly Gly Arg Val His Arg Ala Pro
                325                 330                 335

Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu Leu
            340                 345                 350

Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu Pro
        355                 360                 365

Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn
    370                 375                 380

Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr Glu
385                 390                 395                 400

Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn Leu
                405                 410                 415

Trp Gly Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Arg Glu
            420                 425                 430

Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr Gly
        435                 440                 445

Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val Ala
    450                 455                 460

Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly His
465                 470                 475                 480

Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp
                485                 490                 495

Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg
            500                 505                 510
```

Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile
515 520 525

Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser Thr
530 535 540

Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg Leu
545 550 555 560

His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser
565 570 575

Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln
580 585 590

Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val Ala
595 600 605

Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser Gly
610 615 620

Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr
625 630 635 640

Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro
645 650 655

Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly
660 665 670

Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu
675 680 685

Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val Arg
690 695 700

Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr Val
705 710 715 720

Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala Arg
725 730 735

Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro
740 745 750

Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys Leu
755 760 765

Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val His
770 775 780

Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val Ala
785 790 795 800

Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val Pro
805 810 815

Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys Glu
820 825 830

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 831 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
1   5   10   15

Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Leu
            20   25   30

Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala Lys
        35   40   45

Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Val Val Val Val Val
    50   55   60

Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu Ala Tyr
65   70   75   80

Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu Ala
                85   90   95

Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Val Arg Leu Glu Val
            100   105   110

Pro Gly Phe Glu Ala Asp Asp Val Leu Ala Thr Leu Ala Lys Arg Ala
        115   120   125

Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg Asp Leu
    130   135   140

Tyr Gln Leu Leu Ser Glu Arg Ile Ala Ile Leu His Pro Glu Gly Tyr
145   150   155   160

Leu Ile Thr Pro Ala Trp Leu Tyr Glu Lys Tyr Gly Leu Arg Pro Glu
                165   170   175

Gln Trp Val Asp Tyr Arg Ala Leu Ala Gly Asp Pro Ser Asp Asn Ile
            180   185   190

Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Gln Arg Leu Ile Arg
        195   200   205

Glu Trp Gly Ser Leu Glu Asn Leu Phe Gln His Leu Asp Gln Val Lys
    210   215   220

Pro Ser Leu Arg Glu Lys Leu Gln Ala Gly Met Glu Ala Leu Ala Leu
225   230   235   240

Ser Arg Lys Leu Ser Gln Val His Thr Asp Leu Pro Leu Glu Val Asp
                245   250   255

Phe Gly Arg Arg Arg Thr Pro Asn Leu Glu Gly Leu Arg Ala Phe Leu
            260   265   270

Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu Leu Glu
        275   280   285

Gly Pro Lys Ala Ala Glu Glu Ala Pro Trp Pro Pro Pro Glu Gly Ala
    290   295   300

Phe Leu Gly Phe Ser Phe Ser Arg Pro Glu Pro Met Trp Ala Glu Leu
305   310   315   320

Leu Ala Leu Ala Gly Ala Trp Glu Gly Arg Leu His Arg Ala Gln Asp
                325   330   335

Pro Leu Arg Gly Leu Arg Asp Leu Lys Gly Val Arg Gly Ile Leu Ala
340 345 350

Lys Asp Leu Ala Val Leu Ala Leu Arg Glu Gly Leu Asp Leu Phe Pro
355 360 365

Glu Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn Thr
370 375 380

Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr Glu Asp
385 390 395 400

Ala Gly Glu Arg Ala Leu Leu Ala Glu Arg Leu Phe Gln Thr Leu Lys
405 410 415

Glu Arg Leu Lys Gly Glu Glu Arg Leu Leu Trp Leu Tyr Glu Glu Val
420 425 430

Glu Lys Pro Leu Ser Arg Val Leu Ala Arg Met Glu Ala Thr Gly Val
435 440 445

Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu Ser Leu Glu Val Glu Ala
450 455 460

Glu Val Arg Gln Leu Glu Glu Glu Val Phe Arg Leu Ala Gly His Pro
465 470 475 480

Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp Glu
485 490 495

Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg Ser
500 505 510

Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile Val
515 520 525

Asp Arg Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Asn Thr Tyr
530 535 540

Ile Asp Pro Leu Pro Ala Leu Val His Pro Lys Thr Gly Arg Leu His
545 550 555 560

Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser Ser
565 570 575

Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln Arg
580 585 590

Ile Arg Arg Ala Phe Val Ala Glu Glu Gly Trp Val Leu Val Val Leu
595 600 605

Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser Gly Asp
610 615 620

Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr Gln
625 630 635 640

Thr Ala Ser Trp Met Phe Gly Val Ser Pro Glu Gly Val Asp Pro Leu
645 650 655

Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly Met
660 665 670

Ser Ala His Arg Leu Ser Gly Glu Leu Ser Ile Pro Tyr Glu Glu Ala
675 680 685

Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Tyr Pro Lys Val Arg Ala
690 695 700

Trp Ile Glu Gly Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr Val Glu
705 710 715 720

Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn Ala Arg Val
725 730 735

Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro Val
740 745 750

Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Arg Leu Phe
755 760 765

Pro Arg Leu Gln Glu Leu Gly Ala Arg Met Leu Leu Gln Val His Asp
770 775 780

Glu Leu Val Leu Glu Ala Pro Lys Asp Arg Ala Glu Arg Val Ala Ala
785 790 795 800

Leu Ala Lys Glu Val Met Glu Gly Val Trp Pro Leu Gln Val Pro Leu
805 810 815

Glu Val Glu Val Gly Leu Gly Glu Asp Trp Leu Ser Ala Lys Glu
820 825 830

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 834 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Glu Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu
1 5 10 15

Val Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly
20 25 30

Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala
35 40 45

Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe
50 55 60

Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu
65 70 75 80

Ala Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
85 90 95

Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu
100 105 110

Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Thr Leu Ala Lys
115 120 125

Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg
130 135 140

```
Asp Leu Tyr Gln Leu Val Ser Asp Arg Val Ala Val Leu His Pro Glu
145                 150                 155                 160

Gly His Leu Ile Thr Pro Glu Trp Leu Trp Glu Lys Tyr Gly Leu Arg
                165                 170                 175

Pro Glu Gln Trp Val Asp Phe Arg Ala Leu Val Gly Asp Pro Ser Asp
            180                 185                 190

Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Leu Lys Leu
        195                 200                 205

Leu Lys Glu Trp Gly Ser Leu Glu Asn Leu Leu Lys Asn Leu Asp Arg
    210                 215                 220

Val Lys Pro Glu Asn Val Arg Glu Lys Ile Lys Ala His Leu Glu Asp
225                 230                 235                 240

Leu Arg Leu Ser Leu Glu Leu Ser Arg Val Arg Thr Asp Leu Pro Leu
                245                 250                 255

Glu Val Asp Leu Ala Gln Gly Arg Glu Pro Asp Arg Glu Gly Leu Arg
            260                 265                 270

Ala Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly
        275                 280                 285

Leu Leu Glu Ala Pro Ala Pro Leu Glu Glu Ala Pro Trp Pro Pro Pro
    290                 295                 300

Glu Gly Ala Phe Val Gly Phe Val Leu Ser Arg Pro Glu Pro Met Trp
305                 310                 315                 320

Ala Glu Leu Lys Ala Leu Ala Ala Cys Arg Asp Gly Arg Val His Arg
                325                 330                 335

Ala Ala Asp Pro Leu Ala Gly Leu Lys Asp Leu Lys Glu Val Arg Gly
            340                 345                 350

Leu Leu Ala Lys Asp Leu Ala Val Leu Ala Ser Arg Glu Gly Leu Asp
        355                 360                 365

Leu Val Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro
    370                 375                 380

Ser Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp
385                 390                 395                 400

Thr Glu Asp Ala Ala His Arg Ala Leu Leu Ser Glu Arg Leu His Arg
                405                 410                 415

Asn Leu Leu Lys Arg Leu Glu Gly Glu Glu Lys Leu Leu Trp Leu Tyr
            420                 425                 430

His Glu Val Glu Lys Pro Leu Ser Arg Val Leu Ala His Met Glu Ala
        435                 440                 445

Thr Gly Val Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu Ser Leu Glu
    450                 455                 460

Leu Ala Glu Glu Ile Arg Arg Leu Glu Glu Glu Val Phe Arg Leu Ala
465                 470                 475                 480

Gly His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu
                485                 490                 495
```

Phe Asp Glu Leu Arg Leu Pro Ala Leu Gly Lys Thr Gln Lys Thr Gly
500 505 510

Lys Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His
515 520 525

Pro Ile Val Glu Lys Ile Leu Gln His Arg Glu Leu Thr Lys Leu Lys
530 535 540

Asn Thr Tyr Val Asp Pro Leu Pro Ser Leu Val His Pro Arg Thr Gly
545 550 555 560

Arg Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu
565 570 575

Ser Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu
580 585 590

Gly Gln Arg Ile Arg Arg Ala Phe Val Ala Glu Ala Gly Trp Ala Leu
595 600 605

Val Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu
610 615 620

Ser Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Lys Asp Ile
625 630 635 640

His Thr Gln Thr Ala Ser Trp Met Phe Gly Val Pro Pro Glu Ala Val
645 650 655

Asp Pro Leu Met Arg Arg Ala Ala Lys Thr Val Asn Phe Gly Val Leu
660 665 670

Tyr Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr
675 680 685

Glu Glu Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys
690 695 700

Val Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Lys Arg Gly
705 710 715 720

Tyr Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn
725 730 735

Ala Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn
740 745 750

Met Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val
755 760 765

Lys Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln
770 775 780

Val His Asp Glu Leu Leu Leu Glu Ala Pro Gln Ala Arg Ala Glu Glu
785 790 795 800

Val Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala
805 810 815

Val Pro Leu Glu Val Glu Val Gly Met Gly Glu Asp Trp Leu Ser Ala
820 825 830

Lys Gly

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2502 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATGNNGGCGA TGCTTCCCCT CTTTGAGCCC AAAGGCCGGG TCCTCCTGGT GGACGGCCAC 60
CACCTGGCCT ACCGCACCTT CTTCGCCCTG AAGGGCCTCA CCACCAGCCG GGGCGAACCG 120
GTGCAGGCGG TCTACGGCTT CGCCAAGAGC CTCCTCAAGG CCCTGAAGGA GGACGGGGAC 180
NNGGCGGTGN TCGTGGTCTT TGACGCCAAG GCCCCCTCCT TCCGCCACGA GGCCTACGAG 240
GCCTACAAGG CGGGCCGGGC CCCCACCCCG GAGGACTTTC CCCGGCAGCT CGCCCTCATC 300
AAGGAGCTGG TGGACCTCCT GGGGCTTGCG CGCCTCGAGG TCCCCGGCTA CGAGGCGGAC 360
GACGTNCTGG CCACCCTGGC CAAGAAGGCG GAAAAGGAGG GGTACGAGGT GCGCATCCTC 420
ACCGCCGACC GCGACCTCTA CCAGCTCCTT TCCGACCGCA TCGCCGTCCT CCACCCCGAG 480
GGGTACCTCA TCACCCCGGC GTGGCTTTGG GAGAAGTACG GCCTGAGGCC GGAGCAGTGG 540
GTGGACTACC GGGCCCTGGC GGGGGACCCC TCCGACAACC TCCCCGGGGT CAAGGGCATC 600
GGGGAGAAGA CCGCCCNGAA GCTCCTCNAG GAGTGGGGGA GCCTGGAAAA CCTCCTCAAG 660
AACCTGGACC GGGTGAAGCC CGCCNTCCGG GAGAAGATCC AGGCCCACAT GGANGACCTG 720
ANGCTCTCCT GGGAGCTNTC CCAGGTGCGC ACCGACCTGC CCCTGGAGGT GGACTTCGCC 780
AAGNGGCGGG AGCCCGACCG GGAGGGGCTT AGGGCCTTTC TGGAGAGGCT GGAGTTTGGC 840
AGCCTCCTCC ACGAGTTCGG CCTCCTGGAG GGCCCCAAGG CCCTGGAGGA GGCCCCCTGG 900
CCCCCGCCGG AAGGGGCCTT CGTGGGCTTT GTCCTTTCCC GCCCCGAGCC CATGTGGGCC 960
GAGCTTCTGG CCCTGGCCGC CGCCAGGGAG GGCCGGGTCC ACCGGGCACC AGACCCCTTT 1020
ANGGGCCTNA GGGACCTNAA GGAGGTGCGG GGNCTCCTCG CCAAGGACCT GGCCGTTTTG 1080
GCCCTGAGGG AGGGCCTNGA CCTCNTGCCC GGGGACGACC CCATGCTCCT CGCCTACCTC 1140
CTGGACCCCT CCAACACCAC CCCCGAGGGG GTGGCCCGGC GCTACGGGGG GGAGTGGACG 1200
GAGGANGCGG GGGAGCGGGC CCTCCTNTCC GAGAGGCTCT TCCNGAACCT NNNGCAGCGC 1260
CTTGAGGGGG AGGAGAGGCT CCTTTGGCTT TACCAGGAGG TGGAGAAGCC CCTTTCCCGG 1320
GTCCTGGCCC ACATGGAGGC CACGGGGGTN CGGCTGGACG TGGCCTACCT CCAGGCCCTN 1380
TCCCTGGAGG TGGCGGAGGA GATCCGCCGC CTCGAGGAGG AGGTCTTCCG CCTGGCCGGC 1440
CACCCCTTCA ACCTCAACTC CCGGGACCAG CTGGAAAGGG TGCTCTTTGA CGAGCTNGGG 1500
CTTCCCGCCA TCGGCAAGAC GGAGAAGACN GGCAAGCGCT CCACCAGCGC CGCCGTGCTG 1560
GAGGCCCTNC GNGAGGCCCA CCCCATCGTG GAGAAGATCC TGCAGTACCG GGAGCTCACC 1620
AAGCTCAAGA ACACCTACAT NGACCCCCTG CCNGNCCTCG TCCACCCCAG GACGGGCCGC 1680

CTCCACACCC GCTTCAACCA GACGGCCACG GCCACGGGCA GGCTTAGTAG CTCCGACCCC 1740

AACCTGCAGA ACATCCCCGT CCGCACCCCN CTGGGCCAGA GGATCCGCCG GGCCTTCGTG 1800

GCCGAGGAGG GNTGGGTGTT GGTGGCCCTG GACTATAGCC AGATAGAGCT CCGGGTCCTG 1860

GCCCACCTCT CCGGGGACGA GAACCTGATC CGGGTCTTCC AGGAGGGGAG GGACATCCAC 1920

ACCCAGACCG CCAGCTGGAT GTTCGGCGTC CCCCCGGAGG CCGTGGACCC CCTGATGCGC 1980

CGGGCGGCCA AGACCATCAA CTTCGGGGTC CTCTACGGCA TGTCCGCCCA CCGCCTCTCC 2040

CAGGAGCTTG CCATCCCCTA CGAGGAGGCG GTGGCCTTCA TTGAGCGCTA CTTCCAGAGC 2100

TTCCCCAAGG TGCGGGCCTG GATTGAGAAG ACCCTGGAGG AGGGCAGGAG GCGGGGGTAC 2160

GTGGAGACCC TCTTCGGCCG CCGGCGCTAC GTGCCCGACC TCAACGCCCG GGTGAAGAGC 2220

GTGCGGGAGG CGGCGGAGCG CATGGCCTTC AACATGCCCG TCCAGGGCAC CGCCGCCGAC 2280

CTCATGAAGC TGGCCATGGT GAAGCTCTTC CCCCGGCTNC AGGAAATGGG GGCCAGGATG 2340

CTCCTNCAGG TCCACGACGA GCTGGTCCTC GAGGCCCCCA AAGAGCGGGC GGAGGNGGTG 2400

GCCGCTTTGG CCAAGGAGGT CATGGAGGGG GTCTATCCCC TGGCCGTGCC CCTGGAGGTG 2460

GAGGTGGGGA TGGGGGAGGA CTGGCTCTCC GCCAAGGAGT AG 2502

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 833 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Met Xaa Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu
1               5                   10                  15

Val Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly
            20                  25                  30

Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala
        35                  40                  45

Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Xaa Val
    50                  55                  60

Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu Ala
65                  70                  75                  80

Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu
                85                  90                  95

Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Xaa Arg Leu Glu
            100                 105                 110

Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Thr Leu Ala Lys Lys
        115                 120                 125

Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg Asp
    130                 135                 140
```

Leu Tyr Gln Leu Leu Ser Asp Arg Ile Ala Val Leu His Pro Glu Gly
145 150 155 160

Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg Pro
165 170 175

Glu Gln Trp Val Asp Tyr Arg Ala Leu Xaa Gly Asp Pro Ser Asp Asn
180 185 190

Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Xaa Lys Leu Leu
195 200 205

Xaa Glu Trp Gly Ser Leu Glu Asn Leu Leu Lys Asn Leu Asp Arg Val
210 215 220

Lys Pro Xaa Xaa Arg Glu Lys Ile Xaa Ala His Met Glu Asp Leu Xaa
225 230 235 240

Leu Ser Xaa Xaa Leu Ser Xaa Val Arg Thr Asp Leu Pro Leu Glu Val
245 250 255

Asp Phe Ala Xaa Arg Arg Glu Pro Asp Arg Glu Gly Leu Arg Ala Phe
260 265 270

Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu Leu
275 280 285

Glu Xaa Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu Gly
290 295 300

Ala Phe Val Gly Phe Val Leu Ser Arg Pro Glu Pro Met Trp Ala Glu
305 310 315 320

Leu Leu Ala Leu Ala Ala Ala Arg Xaa Gly Arg Val His Arg Ala Xaa
325 330 335

Asp Pro Leu Xaa Gly Leu Arg Asp Leu Lys Glu Val Arg Gly Leu Leu
340 345 350

Ala Lys Asp Leu Ala Val Leu Ala Leu Arg Glu Gly Leu Asp Leu Xaa
355 360 365

Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn
370 375 380

Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr Glu
385 390 395 400

Asp Ala Gly Glu Arg Ala Leu Leu Ser Glu Arg Leu Phe Xaa Asn Leu
405 410 415

Xaa Xaa Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Xaa Glu
420 425 430

Val Glu Lys Pro Leu Ser Arg Val Leu Ala His Met Glu Ala Thr Gly
435 440 445

Val Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu Ser Leu Glu Val Ala
450 455 460

Glu Glu Ile Arg Arg Leu Glu Glu Glu Val Phe Arg Leu Ala Gly His
465 470 475 480

Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp
485 490 495

```
Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg
            500                 505                 510

Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile
        515                 520                 525

Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Asn Thr
    530                 535                 540

Tyr Ile Asp Pro Leu Pro Xaa Leu Val His Pro Arg Thr Gly Arg Leu
545                 550                 555                 560

His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser
                565                 570                 575

Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln
            580                 585                 590

Arg Ile Arg Arg Ala Phe Val Ala Glu Glu Gly Trp Xaa Leu Val Ala
        595                 600                 605

Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser Gly
    610                 615                 620

Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr
625                 630                 635                 640

Gln Thr Ala Ser Trp Met Phe Gly Val Pro Pro Glu Ala Val Asp Pro
                645                 650                 655

Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly
            660                 665                 670

Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu
        675                 680                 685

Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val Arg
    690                 695                 700

Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr Val
705                 710                 715                 720

Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn Ala Arg
                725                 730                 735

Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro
            740                 745                 750

Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys Leu
        755                 760                 765

Phe Pro Arg Leu Xaa Glu Met Gly Ala Arg Met Leu Leu Gln Val His
    770                 775                 780

Asp Glu Leu Val Leu Glu Ala Pro Lys Xaa Arg Ala Glu Xaa Val Ala
785                 790                 795                 800

Ala Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val Pro
                805                 810                 815

Leu Glu Val Glu Val Gly Xaa Gly Glu Asp Trp Leu Ser Ala Lys Glu
            820                 825                 830

Xaa
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1647 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
ATGAATTCGG GGATGCTGCC CCTCTTTGAG CCCAAGGGCC GGGTCCTCCT GGTGGACGGC     60
CACCACCTGG CCTACCGCAC CTTCCACGCC CTGAAGGGCC TCACCACCAG CCGGGGGGAG    120
CCGGTGCAGG CGGTCTACGG CTTCGCCAAG AGCCTCCTCA AGGCCCTCAA GGAGGACGGG    180
GACGCGGTGA TCGTGGTCTT TGACGCCAAG GCCCCCTCCT TCCGCCACGA GGCCTACGGG    240
GGGTACAAGG CGGGCCGGGC CCCCACGCCG GAGGACTTTC CCCGGCAACT CGCCCTCATC    300
AAGGAGCTGG TGGACCTCCT GGGGCTGGCG CGCCTCGAGG TCCCGGGCTA CGAGGCGGAC    360
GACGTCCTGG CCAGCCTGGC CAAGAAGGCG GAAAAGGAGG GCTACGAGGT CCGCATCCTC    420
ACCGCCGACA AAGACCTTTA CCAGCTCCTT TCCGACCGCA TCCACGTCCT CCACCCCGAG    480
GGGTACCTCA TCACCCCGGC CTGGCTTTGG GAAAAGTACG GCCTGAGGCC CGACCAGTGG    540
GCCGACTACC GGGCCCTGAC CGGGGACGAG TCCGACAACC TTCCCGGGGT CAAGGGCATC    600
GGGGAGAAGA CGGCGAGGAA GCTTCTGGAG GAGTGGGGGA GCCTGGAAGC CCTCCTCAAG    660
AACCTGGACC GGCTGAAGCC CGCCATCCGG GAGAAGATCC TGGCCCACAT GGACGATCTG    720
AAGCTCTCCT GGGACCTGGC CAAGGTGCGC ACCGACCTGC CCCTGGAGGT GGACTTCGCC    780
AAAAGGCGGG AGCCCGACCG GGAGAGGCTT AGGGCCTTTC TGGAGAGGCT TGAGTTTGGC    840
AGCCTCCTCC ACGAGTTCGG CCTTCTGGAA AGCCCCAAGG CCCTGGAGGA GGCCCCCTGG    900
CCCCCGCCGG AAGGGGCCTT CGTGGGCTTT GTGCTTTCCC GCAAGGAGCC CATGTGGGCC    960
GATCTTCTGG CCCTGGCCGC CGCCAGGGGG GGCCGGGTCC ACCGGGCCCC CGAGCCTTAT   1020
AAAGCCCTCA GGGACCTGAA GGAGGCGCGG GGGCTTCTCG CCAAAGACCT GAGCGTTCTG   1080
GCCCTGAGGG AAGGCCTTGG CCTCCCGCCC GGCGACGACC CCATGCTCCT CGCCTACCTC   1140
CTGGACCCTT CCAACACCAC CCCCGAGGGG GTGGCCCGGC GCTACGGCGG GGAGTGGACG   1200
GAGGAGGCGG GGGAGCGGGC CGCCCTTTCC GAGAGGCTCT TCGCCAACCT GTGGGGGAGG   1260
CTTGAGGGGG AGGAGAGGCT CCTTTGGCTT TACCGGGAGG TGGAGAGGCC CCTTTCCGCT   1320
GTCCTGGCCC ACATGGAGGC CACGGGGGTG CGCCTGGACG TGGCCTATCT CAGGGCCTTG   1380
TCCCTGGAGG TGGCCGGGGA GATCGCCCGC CTCGAGGCCG AGGTCTTCCG CCTGGCCGGC   1440
CACCCCTTCA ACCTCAACTC CCGGGACCAG CTGGAAAGGG TCCTCTTTGA CGAGCTAGGG   1500
CTTCCCGCCA TCGGCAAGAC GGAGAAGACC GGCAAGCGCT CCACCAGCGC CGCCGTCCTG   1560
GAGGCCCTCC GCGAGGCCCA CCCCATCGTG GAGAAGATCC TGCAGGCATG CAAGCTTGGC   1620
ACTGGCCGTC GTTTTACAAC GTCGTGA                                        1647
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2088 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATGAATTCGG GGATGCTGCC CCTCTTTGAG CCCAAGGGCC GGGTCCTCCT GGTGGACGGC 60

CACCACCTGG CCTACCGCAC CTTCCACGCC CTGAAGGGCC TCACCACCAG CCGGGGGGAG 120

CCGGTGCAGG CGGTCTACGG CTTCGCCAAG AGCCTCCTCA AGGCCCTCAA GGAGGACGGG 180

GACGCGGTGA TCGTGGTCTT TGACGCCAAG GCCCCCTCCT TCCGCCACGA GGCCTACGGG 240

GGGTACAAGG CGGGCCGGGC CCCCACGCCG GAGGACTTTC CCCGGCAACT CGCCCTCATC 300

AAGGAGCTGG TGGACCTCCT GGGGCTGGCG CGCCTCGAGG TCCCGGGCTA CGAGGCGGAC 360

GACGTCCTGG CCAGCCTGGC CAAGAAGGCG GAAAAGGAGG GCTACGAGGT CCGCATCCTC 420

ACCGCCGACA AAGACCTTTA CCAGCTCCTT TCCGACCGCA TCCACGTCCT CCACCCCGAG 480

GGGTACCTCA TCACCCCGGC CTGGCTTTGG GAAAAGTACG GCCTGAGGCC CGACCAGTGG 540

GCCGACTACC GGGCCCTGAC CGGGGACGAG TCCGACAACC TTCCCGGGGT CAAGGGCATC 600

GGGGAGAAGA CGGCGAGGAA GCTTCTGGAG GAGTGGGGGA GCCTGGAAGC CCTCCTCAAG 660

AACCTGGACC GGCTGAAGCC CGCCATCCGG GAGAAGATCC TGGCCCACAT GGACGATCTG 720

AAGCTCTCCT GGGACCTGGC CAAGGTGCGC ACCGACCTGC CCCTGGAGGT GGACTTCGCC 780

AAAAGGCGGG AGCCCGACCG GGAGAGGCTT AGGGCCTTTC TGGAGAGGCT TGAGTTTGGC 840

AGCCTCCTCC ACGAGTTCGG CCTTCTGGAA AGCCCCAAGG CCCTGGAGGA GGCCCCCTGG 900

CCCCCGCCGG AAGGGGCCTT CGTGGGCTTT GTGCTTTCCC GCAAGGAGCC CATGTGGGCC 960

GATCTTCTGG CCCTGGCCGC CGCCAGGGGG GGCCGGGTCC ACCGGGCCCC CGAGCCTTAT 1020

AAAGCCCTCA GGGACCTGAA GGAGGCGCGG GGGCTTCTCG CCAAAGACCT GAGCGTTCTG 1080

GCCCTGAGGG AAGGCCTTGG CCTCCCGCCC GGCGACGACC CCATGCTCCT CGCCTACCTC 1140

CTGGACCCTT CCAACACCAC CCCCGAGGGG GTGGCCCGGC GCTACGGCGG GGAGTGGACG 1200

GAGGAGGCGG GGGAGCGGGC CGCCCTTTCC GAGAGGCTCT TCGCCAACCT GTGGGGGAGG 1260

CTTGAGGGGG AGGAGAGGCT CCTTTGGCTT TACCGGGAGG TGGAGAGGCC CCTTTCCGCT 1320

GTCCTGGCCC ACATGGAGGC CACGGGGGTG CGCCTGGACG TGGCCTATCT CAGGGCCTTG 1380

TCCCTGGAGG TGGCCGGGGA GATCGCCCGC CTCGAGGCCG AGGTCTTCCG CCTGGCCGGC 1440

CACCCCTTCA ACCTCAACTC CCGGGACCAG CTGGAAAGGG TCCTCTTTGA CGAGCTAGGG 1500

CTTCCCGCCA TCGGCAAGAC GGAGAAGACC GGCAAGCGCT CCACCAGCGC CGCCGTCCTG 1560

GAGGCCCTCC GCGAGGCCCA CCCCATCGTG GAGAAGATCC TGCAGTACCG GGAGCTCACC 1620

AAGCTGAAGA GCACCTACAT TGACCCCTTG CCGGACCTCA TCCACCCCAG GACGGGCCGC 1680

CTCCACACCC GCTTCAACCA GACGGCCACG GCCACGGGCA GGCTAAGTAG CTCCGATCCC 1740

AACCTCCAGA ACATCCCCGT CCGCACCCCG CTTGGGCAGA GGATCCGCCG GGCCTTCATC 1800

GCCGAGGAGG GGTGGCTATT GGTGGCCCTG GACTATAGCC AGATAGAGCT CAGGGTGCTG 1860

GCCCACCTCT CCGGCGACGA GAACCTGATC CGGGTCTTCC AGGAGGGGCG GGACATCCAC 1920

ACGGAGACCG CCAGCTGGAT GTTCGGCGTC CCCCGGGAGG CCGTGGACCC CCTGATGCGC 1980

CGGGCGGCCA AGACCATCAA CTTCGGGGTC CTCTACGGCA TGTCGGCCCA CCGCCTCTCC 2040

CAGGAGCTAG CTAGCCATCC CTTACGAGGA GGCCCAGGCC TTCATTGA 2088

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 962 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATGAATTCGG GGATGCTGCC CCTCTTTGAG CCCAAGGGCC GGGTCCTCCT GGTGGACGGC 60

CACCACCTGG CCTACCGCAC CTTCCACGCC CTGAAGGGCC TCACCACCAG CCGGGGGGAG 120

CCGGTGCAGG CGGTCTACGG CTTCGCCAAG AGCCTCCTCA AGGCCCTCAA GGAGGACGGG 180

GACGCGGTGA TCGTGGTCTT TGACGCCAAG GCCCCCTCCT TCCGCCACGA GGCCTACGGG 240

GGGTACAAGG CGGGCCGGGC CCCCACGCCG GAGGACTTTC CCCGGCAACT CGCCCTCATC 300

AAGGAGCTGG TGGACCTCCT GGGGCTGGCG CGCCTCGAGG TCCCGGGCTA CGAGGCGGAC 360

GACGTCCTGG CCAGCCTGGC CAAGAAGGCG GAAAAGGAGG GCTACGAGGT CCGCATCCTC 420

ACCGCCGACA AAGACCTTTA CCAGCTTCTT TCCGACCGCA TCCACGTCCT CCACCCCGAG 480

GGGTACCTCA TCACCCCGGC CTGGCTTTGG GAAAAGTACG GCCTGAGGCC CGACCAGTGG 540

GCCGACTACC GGGCCCTGAC CGGGGACGAG TCCGACAACC TTCCCGGGGT CAAGGGCATC 600

GGGGAGAAGA CGGCGAGGAA GCTTCTGGAG GAGTGGGGGA GCCTGGAAGC CCTCCTCAAG 660

AACCTGGACC GGCTGAAGCC CGCCATCCGG GAGAAGATCC TGGCCCACAT GGACGATCTG 720

AAGCTCTCCT GGGACCTGGC CAAGGTGCGC ACCGACCTGC CCCTGGAGGT GGACTTCGCC 780

AAAAGGCGGG AGCCCGACCG GGAGAGGCTT AGGGCCTTTC TGGAGAGGCT TGAGTTTGGC 840

AGCCTCCTCC ACGAGTTCGG CCTTCTGGAA AGCCCCAAGT CATGGAGGGG GTGTATCCCC 900

TGGCCGTGCC CCTGGAGGTG GAGGTGGGGA TAGGGGAGGA CTGGCTCTCC GCCAAGGAGT 960

GA 962

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1600 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
ATGGAATTCG GGGATGCTGC CCCTCTTTGA GCCCAAGGGC CGGGTCCTCC TGGTGGACGG     60
CCACCACCTG GCCTACCGCA CCTTCCACGC CCTGAAGGGC CTCACCACCA GCCGGGGGGA    120
GCCGGTGCAG GCGGTCTACG GCTTCGCCAA GAGCCTCCTC AAGGCCCTCA AGGAGGACGG    180
GGACGCGGTG ATCGTGGTCT TTGACGCCAA GGCCCCCTCC TTCCGCCACG AGGCCTACGG    240
GGGGTACAAG GCGGGCCGGG CCCCCACGCC GGAGGACTTT CCCCGGCAAC TCGCCCTCAT    300
CAAGGAGCTG GTGGACCTCC TGGGGCTGGC GCGCCTCGAG GTCCCGGGCT ACGAGGCGGA    360
CGACGTCCTG GCCAGCCTGG CCAAGAAGGC GGAAAAGGAG GGCTACGAGG TCCGCATCCT    420
CACCGCCGAC AAAGACCTTT ACCAGCTCCT TTCCGACCGC ATCCACGTCC TCCACCCCGA    480
GGGGTACCTC ATCACCCCGG CCTGGCTTTG GGAAAAGTAC GGCCTGAGGC CCGACCAGTG    540
GGCCGACTAC CGGGCCCTGA CCGGGGACGA GTCCGACAAC CTTCCCGGGG TCAAGGGCAT    600
CGGGGAGAAG ACGGCGAGGA AGCTTCTGGA GGAGTGGGGG AGCCTGGAAG CCCTCCTCAA    660
GAACCTGGAC CGGCTGAAGC CCGCCATCCG GGAGAAGATC CTGGCCCACA TGGACGATCT    720
GAAGCTCTCC TGGGACCTGG CCAAGGTGCG CACCGACCTG CCCCTGGAGG TGGACTTCGC    780
CAAAAGGCGG GAGCCCGACC GGGAGAGGCT TAGGGCCTTT CTGGAGAGGC TTGAGTTTGG    840
CAGCCTCCTC CACGAGTTCG GCCTTCTGGA AAGCCCCAAG ATCCGCCGGG CCTTCATCGC    900
CGAGGAGGGG TGGCTATTGG TGGCCCTGGA CTATAGCCAG ATAGAGCTCA GGGTGCTGGC    960
CCACCTCTCC GGCGACGAGA ACCTGATCCG GGTCTTCCAG GAGGGGCGGG ACATCCACAC   1020
GGAGACCGCC AGCTGGATGT TCGGCGTCCC CCGGGAGGCC GTGGACCCCC TGATGCGCCG   1080
GGCGGCCAAG ACCATCAACT TCGGGGTCCT CTACGGCATG TCGGCCCACC GCCTCTCCCA   1140
GGAGCTAGCC ATCCCTTACG AGGAGGCCCA GGCCTTCATT GAGCGCTACT TTCAGAGCTT   1200
CCCCAAGGTG CGGGCCTGGA TTGAGAAGAC CCTGGAGGAG GGCAGGAGGC GGGGGTACGT   1260
GGAGACCCTC TTCGGCCGCC GCCGCTACGT GCCAGACCTA GAGGCCCGGG TGAAGAGCGT   1320
GCGGGAGGCG GCCGAGCGCA TGGCCTTCAA CATGCCCGTC CGGGGCACCG CCGCCGACCT   1380
CATGAAGCTG GCTATGGTGA AGCTCTTCCC CAGGCTGGAG GAAATGGGGG CCAGGATGCT   1440
CCTTCAGGTC CACGACGAGC TGGTCCTCGA GGCCCCAAAA GAGAGGGCGG AGGCCGTGGC   1500
CCGGCTGGCC AAGGAGGTCA TGGAGGGGGT GTATCCCCTG GCCGTGCCCC TGGAGGTGGA   1560
GGTGGGGATA GGGGAGGACT GGCTCTCCGC CAAGGAGTGA                          1600
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CACGAATTCG GGGATGCTGC CCCTCTTTGA GCCCAA 36

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GTGAGATCTA TCACTCCTTG GCGGAGAGCC AGTC 34

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 91 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TAATACGACT CACTATAGGG AGACCGGAAT TCGAGCTCGC CCGGGCGAGC TCGAATTCCG 60

TGTATTCTAT AGTGTCACCT AAATCGAATT C 91

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TAATACGACT CACTATAGGG 20

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GAATTCGATT TAGGTGACAC TATAGAA 27

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GTAATCATGG TCATAGCTGG TAGCTTGCTA C 31

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGATCCTCTA GAGTCGACCT GCAGGCATGC CTACCTTGGT AG 42

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GGATCCTCTA GAGTCGACCT GCAGGCATGC 30

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2502 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ATGAATTCGG GGATGCTGCC CCTCTTTGAG CCCAAGGGCC GGGTCCTCCT GGTGGACGGC 60

CACCACCTGG CCTACCGCAC CTTCCACGCC CTGAAGGGCC TCACCACCAG CCGGGGGGAG 120

CCGGTGCAGG CGGTCTACGG CTTCGCCAAG AGCCTCCTCA AGGCCCTCAA GGAGGACGGG 180

GACGCGGTGA TCGTGGTCTT TGACGCCAAG GCCCCCTCCT TCCGCCACGA GGCCTACGGG 240

GGGTACAAGG CGGGCCGGGC CCCCACGCCG GAGGACTTTC CCCGGCAACT CGCCCTCATC 300

AAGGAGCTGG TGGACCTCCT GGGGCTGGCG CGCCTCGAGG TCCCGGGCTA CGAGGCGGAC 360

GACGTCCTGG CCAGCCTGGC CAAGAAGGCG GAAAAGGAGG GCTACGAGGT CCGCATCCTC 420

ACCGCCGACA AAGACCTTTA CCAGCTCCTT TCCGACCGCA TCCACGTCCT CCACCCCGAG 480
GGGTACCTCA TCACCCCGGC CTGGCTTTGG GAAAAGTACG GCCTGAGGCC CGACCAGTGG 540
GCCGACTACC GGGCCCTGAC CGGGGACGAG TCCGACAACC TTCCCGGGGT CAAGGGCATC 600
GGGGAGAAGA CGGCGAGGAA GCTTCTGGAG GAGTGGGGGA GCCTGGAAGC CCTCCTCAAG 660
AACCTGGACC GGCTGAAGCC CGCCATCCGG GAGAAGATCC TGGCCCACAT GGACGATCTG 720
AAGCTCTCCT GGGACCTGGC CAAGGTGCGC ACCGACCTGC CCCTGGAGGT GGACTTCGCC 780
AAAAGGCGGG AGCCCGACCG GGAGAGGCTT AGGGCCTTTC TGGAGAGGCT TGAGTTTGGC 840
AGCCTCCTCC ACGAGTTCGG CCTTCTGGAA AGCCCCAAGG CCCTGGAGGA GGCCCCCTGG 900
CCCCCGCCGG AAGGGGCCTT CGTGGGCTTT GTGCTTTCCC GCAAGGAGCC CATGTGGGCC 960
GATCTTCTGG CCCTGGCCGC CGCCAGGGGG GGCCGGGTCC ACCGGGCCCC CGAGCCTTAT 1020
AAAGCCCTCA GGGACCTGAA GGAGGCGCGG GGGCTTCTCG CCAAAGACCT GAGCGTTCTG 1080
GCCCTGAGGG AAGGCCTTGG CCTCCCGCCC GGCGACGACC CCATGCTCCT CGCCTACCTC 1140
CTGGACCCTT CCAACACCAC CCCCGAGGGG GTGGCCCGGC GCTACGGCGG GGAGTGGACG 1200
GAGGAGGCGG GGGAGCGGGC CGCCCTTTCC GAGAGGCTCT TCGCCAACCT GTGGGGGAGG 1260
CTTGAGGGGG AGGAGAGGCT CCTTTGGCTT TACCGGGAGG TGGAGAGGCC CCTTTCCGCT 1320
GTCCTGGCCC ACATGGAGGC CACGGGGGTG CGCCTGGACG TGGCCTATCT CAGGGCCTTG 1380
TCCCTGGAGG TGGCCGGGGA GATCGCCCGC CTCGAGGCCG AGGTCTTCCG CCTGGCCGGC 1440
CACCCCTTCA ACCTCAACTC CCGGGACCAG CTGGAAAGGG TCCTCTTTGA CGAGCTAGGG 1500
CTTCCCGCCA TCGGCAAGAC GGAGAAGACC GGCAAGCGCT CCACCAGCGC CGCCGTCCTG 1560
GAGGCCCTCC GCGAGGCCCA CCCCATCGTG GAGAAGATCC TGCAGTACCG GGAGCTCACC 1620
AAGCTGAAGA GCACCTACAT TGACCCCTTG CCGGACCTCA TCCACCCCAG GACGGGCCGC 1680
CTCCACACCC GCTTCAACCA GACGGCCACG GCCACGGGCA GGCTAAGTAG CTCCGATCCC 1740
AACCTCCAGA ACATCCCCGT CCGCACCCCG CTTGGGCAGA GGATCCGCCG GGCCTTCATC 1800
GCCGAGGAGG GGTGGCTATT GGTGGCCCTG GACTATAGCC AGATAGAGCT CAGGGTGCTG 1860
GCCCACCTCT CCGGCGACGA GAACCTGATC CGGGTCTTCC AGGAGGGGCG GGACATCCAC 1920
ACGGAGACCG CCAGCTGGAT GTTCGGCGTC CCCCGGGAGG CCGTGGACCC CCTGATGCGC 1980
CGGGCGGCCA AGACCATCAA CTTCGGGGTC CTCTACGGCA TGTCGGCCCA CCGCCTCTCC 2040
CAGGAGCTAG CCATCCCTTA CGAGGAGGCC CAGGCCTTCA TTGAGCGCTA CTTTCAGAGC 2100
TTCCCCAAGG TGCGGGCCTG GATTGAGAAG ACCCTGGAGG AGGGCAGGAG GCGGGGGTAC 2160
GTGGAGACCC TCTTCGGCCG CCGCCGCTAC GTGCCAGACC TAGAGGCCCG GGTGAAGAGC 2220
GTGCGGGAGG CGGCCGAGCG CATGGCCTTC AACATGCCCG TCCGGGGCAC CGCCGCCGAC 2280
CTCATGAAGC TGGCTATGGT GAAGCTCTTC CCCAGGCTGG AGGAAATGGG GGCCAGGATG 2340
CTCCTTCAGG TCCACGACGA GCTGGTCCTC GAGGCCCCAA AAGAGAGGGC GGAGGCCGTG 2400

GCCCGGCTGG CCAAGGAGGT CATGGAGGGG GTGTATCCCC TGGCCGTGCC CCTGGAGGTG 2460

GAGGTGGGGA TAGGGGAGGA CTGGCTCTCC GCCAAGGAGT GA 2502

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GATTTAGGTG ACACTATAG 19

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

ACACAGGTAC CACATGGTAC AAGAGGCAAG AGAGACGACA CAGCAGAAAC 50

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Ile Asn Ser
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 969 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ATGGCTAGCA TGACTGGTGG ACAGCAAATG GGTCGGATCA ATTCGGGGAT GCTGCCCCTC 60

TTTGAGCCCA AGGGCCGGGT CCTCCTGGTG GACGGCCACC ACCTGGCCTA CCGCACCTTC 120

CACGCCCTGA AGGGCCTCAC CACCAGCCGG GGGGAGCCGG TGCAGGCGGT CTACGGCTTC 180

GCCAAGAGCC TCCTCAAGGC CCTCAAGGAG GACGGGGACG CGGTGATCGT GGTCTTTGAC 240

GCCAAGGCCC CCTCCTTCCG CCACGAGGCC TACGGGGGGT ACAAGGCGGG CCGGGCCCCC 300
ACGCCGGAGG ACTTTCCCCG GCAACTCGCC CTCATCAAGG AGCTGGTGGA CCTCCTGGGG 360
CTGGCGCGCC TCGAGGTCCC GGGCTACGAG GCGGACGACG TCCTGGCCAG CCTGGCCAAG 420
AAGGCGGAAA AGGAGGGCTA CGAGGTCCGC ATCCTCACCG CCGACAAAGA CCTTTACCAG 480
CTTCTTTCCG ACCGCATCCA CGTCCTCCAC CCCGAGGGGT ACCTCATCAC CCCGGCCTGG 540
CTTTGGGAAA AGTACGGCCT GAGGCCCGAC CAGTGGGCCG ACTACCGGGC CCTGACCGGG 600
GACGAGTCCG ACAACCTTCC CGGGGTCAAG GGCATCGGGG AGAAGACGGC GAGGAAGCTT 660
CTGGAGGAGT GGGGGAGCCT GGAAGCCCTC CTCAAGAACC TGGACCGGCT GAAGCCCGCC 720
ATCCGGGAGA AGATCCTGGC CCACATGGAC GATCTGAAGC TCTCCTGGGA CCTGGCCAAG 780
GTGCGCACCG ACCTGCCCCT GGAGGTGGAC TTCGCCAAAA GGCGGGAGCC CGACCGGGAG 840
AGGCTTAGGG CCTTTCTGGA GAGGCTTGAG TTTGGCAGCC TCCTCCACGA GTTCGGCCTT 900
CTGGAAAGCC CCAAGTCATG GAGGGGGTGT ATCCCCTGGC CGTGCCCCTG GAGGTGGAGG 960
TGGGGATAG 969

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 948 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

ATGGCTAGCA TGACTGGTGG ACAGCAAATG GGTCGGATCA ATTCGGGGAT GCTGCCCCTC 60
TTTGAGCCCA AGGGCCGGGT CCTCCTGGTG GACGGCCACC ACCTGGCCTA CCGCACCTTC 120
CACGCCCTGA AGGGCCTCAC CACCAGCCGG GGGGAGCCGG TGCAGGCGGT CTACGGCTTC 180
GCCAAGAGCC TCCTCAAGGC CCTCAAGGAG GACGGGGACG CGGTGATCGT GGTCTTTGAC 240
GCCAAGGCCC CCTCCTTCCG CCACGAGGCC TACGGGGGGT ACAAGGCGGG CCGGGCCCCC 300
ACGCCGGAGG ACTTTCCCCG GCAACTCGCC CTCATCAAGG AGCTGGTGGA CCTCCTGGGG 360
CTGGCGCGCC TCGAGGTCCC GGGCTACGAG GCGGACGACG TCCTGGCCAG CCTGGCCAAG 420
AAGGCGGAAA AGGAGGGCTA CGAGGTCCGC ATCCTCACCG CCGACAAAGA CCTTTACCAG 480
CTTCTTTCCG ACCGCATCCA CGTCCTCCAC CCCGAGGGGT ACCTCATCAC CCCGGCCTGG 540
CTTTGGGAAA AGTACGGCCT GAGGCCCGAC CAGTGGGCCG ACTACCGGGC CCTGACCGGG 600
GACGAGTCCG ACAACCTTCC CGGGGTCAAG GGCATCGGGG AGAAGACGGC GAGGAAGCTT 660
CTGGAGGAGT GGGGGAGCCT GGAAGCCCTC CTCAAGAACC TGGACCGGCT GAAGCCCGCC 720
ATCCGGGAGA AGATCCTGGC CCACATGGAC GATCTGAAGC TCTCCTGGGA CCTGGCCAAG 780
GTGCGCACCG ACCTGCCCCT GGAGGTGGAC TTCGCCAAAA GGCGGGAGCC CGACCGGGAG 840

AGGCTTAGGG CCTTTCTGGA GAGGCTTGAG TTTGGCAGCC TCCTCCACGA GTTCGGCCTT 900

CTGGAAAGCC CCAAGGCCGC ACTCGAGCAC CACCACCACC ACCACTGA 948

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 206 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CGCCAGGGTT TTCCCAGTCA CGACGTTGTA AAACGACGGC CAGTGAATTG TAATACGACT 60

CACTATAGGG CGAATTCGAG CTCGGTACCC GGGGATCCTC TAGAGTCGAC CTGCAGGCAT 120

GCAAGCTTGA GTATTCTATA GTGTCACCTA AATAGCTTGG CGTAATCATG GTCATAGCTG 180

TTTCCTGTGT GAAATTGTTA TCCGCT 206

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

AACAGCTATG ACCATGATTA C 21

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GTTCTCTGCT CTCTGGTCGC TGTCTCGCTT GTGAAACAAG CGAGACAGCG TGGTCTCTCG 60

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CGAGAGACCA CGCTG 15

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CCTTTCGCTT TCTTCCCTTC CTTTCTCGCC ACGTTCGCCG GCTTTCCCCG TC 52

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

AGAAAGGAAG GGAAGAAAGC GAAAGG 26

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GACGGGGAAA GCCGGCGAAC G 21

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GAAAGCCGGC GAACGTGGCG 20

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GGCGAACGTG GCGAGAAAGG A 21

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CCTTTCGCTT TCTTCCCTTC CTTTCTCGCC ACGTTCGCCG GC 42

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CCTTTCGCTC TCTTCCCTTC CTTTCTCGCC ACGTTCGCCG GC 42

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 8
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/mod_base= OTHER
/note= "The A residue at this position is 2'-O-methyladenosine."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

AGAAAGGAAG GGAAGAAAGC GAAAGGT 27

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GCCGGCGAAC GTGGCGAGAA AGGA 24

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GGTTTTTCTT TGAGGTTTAG 20

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCGACACTCC ACCATAGAT 19

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CTGTCTTCAC GCAGAAAGC 19

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GCACGGTCTA CGAGACCTC 19

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

TAATACGACT CACTATAGGG 20

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 337 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: RNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GGGAAAGCUU GCAUGCCUGC AGGUCGACUC UAGAGGAUCU ACUAGUCAUA UGGAUUCUGU 60

CUUCACGCAG AAAGCGUCUG GCCAUGGCGU UAGUAUGAGU GUCGUGCAGC CUCCAGGACC 120

CCCCCUCCCG GGAGAGGCAU AGUGGUCUGC GGAACCGGUG AGUACACCGG AAUUGCCAGG 180

ACGACCGGGU CCUUUCUUGG AUAAACCCGC UCAAUGCCUG GAGAUUUGGG CGUGCCCCCG 240

CAAGACUGCU AGCCGAGUAG UGUUGGGUCG CGAAAGGCCU UGUGGUACUG CCUGAUAGGG 300

UGCCUGCGAG UGCCCCGGGA GGUCUCGUAG ACCGUGC 337

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 18
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/note= "The N at this position indicates the presence of a fluorescein dye on an abasic linker."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CCGGTCGTCC TGGCAATNCC 20

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GTTTATCCAA GAAAGGACCC GGTC 24

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CAGGGTGAAG GGAAGAAGAA AGCGAAAGGT 30

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CAGGGGGAAG GGAAGAAGAA AGCGAAAGGT 30

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1..2
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/mod_base= OTHER
/note= "The T residues at positions 1 and 2 are amino modified T residues."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

TTCTTTTCAC CAGCGAGACG GG 22

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

ATTGGGCGCC AGGGTGGTTT TT 22

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

CCCGTCTCGC TGGTGAAAAG AAAAACCACC CTGGCGCCCA ATACGCAAAC CGC 53

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GAATTCGATT TAGGTGACAC TATAGAATAC A 31

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

CCTTTCGCTT TCTTCCCTTC CTTTCTCGCC ACGTTCGCCG GC 42

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GCCGGCGAAC GTGGCGAGAA AGGA 24

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

CAGAAGGAAG GGAAGAAAGC GAAAGG 26

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CAGGGGGAAG GGAAGAAAGC GAAAGG 26

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

CAGGGTACAG GGAAGAAAGC GAAAGG 26

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

GGGAAAGTCC TCGGAGCCGC GCGGGACGAG CGTGGGGGCC CG 42

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 963 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)
(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..960
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

```
ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT CGG ATC AAT TCG GGG        48
Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Ile Asn Ser Gly
  1               5                  10                  15
```

```
ATG CTG CCC CTC TTT GAG CCC AAG GGC CGG GTC CTC CTG GTG GAC GGC      96
Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val Asp Gly
            20                  25                  30

CAC CAC CTG GCC TAC CGC ACC TTC CAC GCC CTG AAG GGC CTC ACC ACC     144
His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys Gly Leu Thr Thr
        35                  40                  45

AGC CGG GGG GAG CCG GTG CAG GCG GTC TAC GGC TTC GCC AAG AGC CTC     192
Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala Lys Ser Leu
    50                  55                  60

CTC AAG GCC CTC AAG GAG GAC GGG GAC GCG GTG ATC GTG GTC TTT GAC     240
Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile Val Val Phe Asp
65                  70                  75                  80

GCC AAG GCC CCC TCC TTC CGC CAC GAG GCC TAC GGG GGG TAC AAG GCG     288
Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly Gly Tyr Lys Ala
                85                  90                  95

GGC CGG GCC CCC ACG CTC GTC CCG CGC GGC TCC GAG GAC TTT CCC CGG     336
Gly Arg Ala Pro Thr Leu Val Pro Arg Gly Ser Glu Asp Phe Pro Arg
            100                 105                 110

CAA CTC GCC CTC ATC AAG GAG CTG GTG GAC CTC CTG GGG CTG GCG CGC     384
Gln Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg
        115                 120                 125

CTC GAG GTC CCG GGC TAC GAG GCG GAC GAC GTC CTG GCC AGC CTG GCC     432
Leu Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala
    130                 135                 140

AAG AAG GCG GAA AAG GAG GGC TAC GAG GTC CGC ATC CTC ACC GCC GAC     480
Lys Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp
145                 150                 155                 160

AAA GAC CTT TAC CAG CTC CTT TCC GAC CGC ATC CAC GTC CTC CAC CCC     528
Lys Asp Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro
                165                 170                 175

GAG GGG TAC CTC ATC ACC CCG GCC TGG CTT TGG GAA AAG TAC GGC CTG     576
Glu Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu
            180                 185                 190

AGG CCC GAC CAG TGG GCC GAC TAC CGG GCC CTG ACC GGG GAC GAG TCC     624
Arg Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser
        195                 200                 205

GAC AAC CTT CCC GGG GTC AAG GGC ATC GGG GAG AAG ACG GCG AGG AAG     672
Asp Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys
    210                 215                 220

CTT CTG GAG GAG TGG GGG AGC CTG GAA GCC CTC CTC AAG AAC CTG GAC     720
Leu Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp
225                 230                 235                 240

CGG CTG AAG CCC GCC ATC CGG GAG AAG ATC CTG GCC CAC ATG GAC GAT     768
Arg Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp
                245                 250                 255

CTG AAG CTC TCC TGG GAC CTG GCC AAG GTG CGC ACC GAC CTG CCC CTG     816
Leu Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu
            260                 265                 270

GAG GTG GAC TTC GCC AAA AGG CGG GAG CCC GAC CGG GAG AGG CTT AGG     864
Glu Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg
        275                 280                 285
```

```
GCC TTT CTG GAG AGG CTT GAG TTT GGC AGC CTC CTC CAC GAG TTC GGC      912
Ala Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly
    290                 295                 300

CTT CTG GAA AGC CCC AAG GCC GCA CTC GAG CAC CAC CAC CAC CAC CAC      960
Leu Leu Glu Ser Pro Lys Ala Ala Leu Glu His His His His His His
305                 310                 315                 320

TGA                                                                  963
```

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 320 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

```
Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Ile Asn Ser Gly
  1               5                  10                  15

Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val Asp Gly
             20                  25                  30

His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys Gly Leu Thr Thr
         35                  40                  45

Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala Lys Ser Leu
     50                  55                  60

Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile Val Val Phe Asp
 65                  70                  75                  80

Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly Gly Tyr Lys Ala
                 85                  90                  95

Gly Arg Ala Pro Thr Leu Val Pro Arg Gly Ser Glu Asp Phe Pro Arg
            100                 105                 110

Gln Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg
        115                 120                 125

Leu Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala
    130                 135                 140

Lys Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp
145                 150                 155                 160

Lys Asp Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro
                165                 170                 175

Glu Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu
            180                 185                 190

Arg Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser
        195                 200                 205

Asp Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys
    210                 215                 220

Leu Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp
225                 230                 235                 240

Arg Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp
                245                 250                 255
```

```
Leu Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu
            260                 265                 270

Glu Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg
        275                 280                 285

Ala Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly
    290                 295                 300

Leu Leu Glu Ser Pro Lys Ala Ala Leu Glu His His His His His His
305                 310                 315                 320
```

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

CGATCTCCTC GGCCACCTCC 20

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GGCGGTGCCC TGGACGGGCA 20

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

CCAGCTCGTT GTGGACCTGA 20

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2505 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..2499

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

```
ATG AAT TCG GGG ATG CTG CCC CTC TTT GAG CCC AAG GGC CGG GTC CTC      48
Met Asn Ser Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu
  1               5                  10                  15

CTG GTG GAC GGC CAC CAC CTG GCC TAC CGC ACC TTC CAC GCC CTG AAG      96
Leu Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys
             20                  25                  30

GGC CTC ACC ACC AGC CGG GGG GAG CCG GTG CAG GCG GTC TAC GGC TTC     144
Gly Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe
         35                  40                  45

GCC AAG AGC CTC CTC AAG GCC CTC AAG GAG GAC GGG GAC GCG GTG ATC     192
Ala Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile
     50                  55                  60

GTG GTC TTT GAC GCC AAG GCC CCC TCC TTC CGC CAC GAG GCC TAC GGG     240
Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly
 65                  70                  75                  80

GGG TAC AAG GCG GGC CGG GCC CCC ACG CCG GAG GAC TTT CCC CGG CAA     288
Gly Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
                 85                  90                  95

CTC GCC CTC ATC AAG GAG CTG GTG GAC CTC CTG GGG CTG GCG CGC CTC     336
Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu
            100                 105                 110

GAG GTC CCG GGC TAC GAG GCG GAC GAC GTC CTG GCC AGC CTG GCC AAG     384
Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys
        115                 120                 125

AAG GCG GAA AAG GAG GGC TAC GAG GTC CGC ATC CTC ACC GCC GAC AAA     432
Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys
    130                 135                 140

GAC CTT TAC CAG CTC CTT TCC GAC CGC ATC CAC GTC CTC CAC CCC GAG     480
Asp Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu
145                 150                 155                 160

GGG TAC CTC ATC ACC CCG GCC TGG CTT TGG GAA AAG TAC GGC CTG AGG     528
Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg
                165                 170                 175

CCC GAC CAG TGG GCC GAC TAC CGG GCC CTG ACC GGG GAC GAG TCC GAC     576
Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp
            180                 185                 190

AAC CTT CCC GGG GTC AAG GGC ATC GGG GAG AAG ACG GCG AGG AAG CTT     624
Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu
        195                 200                 205

CTG GAG GAG TGG GGG AGC CTG GAA GCC CTC CTC AAG AAC CTG GAC CGG     672
Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg
    210                 215                 220

CTG AAG CCC GCC ATC CGG GAG AAG ATC CTG GCC CAC ATG GAC GAT CTG     720
Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu
225                 230                 235                 240
```

```
AAG CTC TCC TGG GAC CTG GCC AAG GTG CGC ACC GAC CTG CCC CTG GAG      768
Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu
                245                 250                 255

GTG GAC TTC GCC AAA AGG CGG GAG CCC GAC CGG GAG AGG CTT AGG GCC      816
Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala
            260                 265                 270

TTT CTG GAG AGG CTT GAG TTT GGC AGC CTC CTC CAC GAG TTC GGC CTT      864
Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu
        275                 280                 285

CTG GAA AGC CCC AAG GCC CTG GAG GAG GCC CCC TGG CCC CCG CCG GAA      912
Leu Glu Ser Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu
    290                 295                 300

GGG GCC TTC GTG GGC TTT GTG CTT TCC CGC AAG GAG CCC ATG TGG GCC      960
Gly Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala
305                 310                 315                 320

GAT CTT CTG GCC CTG GCC GCC GCC AGG GGG GGC CGG GTC CAC CGG GCC     1008
Asp Leu Leu Ala Leu Ala Ala Ala Arg Gly Gly Arg Val His Arg Ala
                325                 330                 335

CCC GAG CCT TAT AAA GCC CTC AGG GAC CTG AAG GAG GCG CGG GGG CTT     1056
Pro Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu
            340                 345                 350

CTC GCC AAA GAC CTG AGC GTT CTG GCC CTG AGG GAA GGC CTT GGC CTC     1104
Leu Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu
        355                 360                 365

CCG CCC GGC GAC GAC CCC ATG CTC CTC GCC TAC CTC CTG GAC CCT TCC     1152
Pro Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser
    370                 375                 380

AAC ACC ACC CCC GAG GGG GTG GCC CGG CGC TAC GGC GGG GAG TGG ACG     1200
Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr
385                 390                 395                 400

GAG GAG GCG GGG GAG CGG GCC GCC CTT TCC GAG AGG CTC TTC GCC AAC     1248
Glu Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn
                405                 410                 415

CTG TGG GGG AGG CTT GAG GGG GAG GAG AGG CTC CTT TGG CTT TAC CGG     1296
Leu Trp Gly Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Arg
            420                 425                 430

GAG GTG GAG AGG CCC CTT TCC GCT GTC CTG GCC CAC ATG GAG GCC ACG     1344
Glu Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr
        435                 440                 445

GGG GTG CGC CTG GAC GTG GCC TAT CTC AGG GCC TTG TCC CTG GAG GTG     1392
Gly Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val
    450                 455                 460

GCC GAG GAG ATC GCC CGC CTC GAG GCC GAG GTC TTC CGC CTG GCC GGC     1440
Ala Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly
465                 470                 475                 480

CAC CCC TTC AAC CTC AAC TCC CGG GAC CAG CTG GAA AGG GTC CTC TTT     1488
His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe
                485                 490                 495

GAC GAG CTA GGG CTT CCC GCC ATC GGC AAG ACG GAG AAG ACC GGC AAG     1536
Asp Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys
            500                 505                 510
```

```
CGC TCC ACC AGC GCC GCC GTC CTG GAG GCC CTC CGC GAG GCC CAC CCC     1584
Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro
        515                 520                 525

ATC GTG GAG AAG ATC CTG CAG TAC CGG GAG CTC ACC AAG CTG AAG AGC     1632
Ile Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser
    530                 535                 540

ACC TAC ATT GAC CCC TTG CCG GAC CTC ATC CAC CCC AGG ACG GGC CGC     1680
Thr Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg
545                 550                 555                 560

CTC CAC ACC CGC TTC AAC CAG ACG GCC ACG GCC ACG GGC AGG CTA AGT     1728
Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser
                565                 570                 575

AGC TCC GAT CCC AAC CTC CAG AAC ATC CCC GTC CGC ACC CCG CTT GGG     1776
Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly
            580                 585                 590

CAG AGG ATC CGC CGG GCC TTC ATC GCC GAG GAG GGG TGG CTA TTG GTG     1824
Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val
        595                 600                 605

GCC CTG GAC TAT AGC CAG ATA GAG CTC AGG GTG CTG GCC CAC CTC TCC     1872
Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser
    610                 615                 620

GGC GAC GAG AAC CTG ATC CGG GTC TTC CAG GAG GGG CGG GAC ATC CAC     1920
Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His
625                 630                 635                 640

ACG GAG ACC GCC AGC TGG ATG TTC GGC GTC CCC CGG GAG GCC GTG GAC     1968
Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp
                645                 650                 655

CCC CTG ATG CGC CGG GCG GCC AAG ACC ATC AAC TTC GGG GTC CTC TAC     2016
Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr
            660                 665                 670

GGC ATG TCG GCC CAC CGC CTC TCC CAG GAG CTA GCC ATC CCT TAC GAG     2064
Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
        675                 680                 685

GAG GCC CAG GCC TTC ATT GAG CGC TAC TTT CAG AGC TTC CCC AAG GTG     2112
Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
    690                 695                 700

CGG GCC TGG ATT GAG AAG ACC CTG GAG GAG GGC AGG AGG CGG GGG TAC     2160
Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr
705                 710                 715                 720

GTG GAG ACC CTC TTC GGC CGC CGC CGC TAC GTG CCA GAC CTA GAG GCC     2208
Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala
                725                 730                 735

CGG GTG AAG AGC GTG CGG GAG GCG GCC GAG CGC ATG GCC TTC AAC ATG     2256
Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
            740                 745                 750

CCC GTC CAG GGC ACC GCC GCC GAC CTC ATG AAG CTG GCT ATG GTG AAG     2304
Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
        755                 760                 765

CTC TTC CCC AGG CTG GAG GAA ATG GGG GCC AGG ATG CTC CTT CAG GTC     2352
Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val
    770                 775                 780
```

```
CAC AAC GAG CTG GTC CTC GAG GCC CCA AAA GAG AGG GCG GAG GCC GTG      2400
His Asn Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val
785                 790                 795                 800

GCC CGG CTG GCC AAG GAG GTC ATG GAG GGG GTG TAT CCC CTG GCC GTG      2448
Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val
                805                 810                 815

CCC CTG GAG GTG GAG GTG GGG ATA GGG GAG GAC TGG CTC TCC GCC AAG      2496
Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
            820                 825                 830

GAG TGATAG                                                           2505
Glu
```

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 833 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

```
Met Asn Ser Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu
  1               5                  10                  15

Leu Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys
             20                  25                  30

Gly Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe
         35                  40                  45

Ala Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile
     50                  55                  60

Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly
 65                  70                  75                  80

Gly Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
                 85                  90                  95

Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu
            100                 105                 110

Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys
        115                 120                 125

Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys
    130                 135                 140

Asp Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu
145                 150                 155                 160

Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg
                165                 170                 175

Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp
            180                 185                 190

Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu
        195                 200                 205

Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg
    210                 215                 220
```

Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu
225 230 235 240

Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu
245 250 255

Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala
260 265 270

Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu
275 280 285

Leu Glu Ser Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu
290 295 300

Gly Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala
305 310 315 320

Asp Leu Leu Ala Leu Ala Ala Ala Arg Gly Gly Arg Val His Arg Ala
325 330 335

Pro Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu
340 345 350

Leu Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu
355 360 365

Pro Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser
370 375 380

Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr
385 390 395 400

Glu Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn
405 410 415

Leu Trp Gly Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Arg
420 425 430

Glu Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr
435 440 445

Gly Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val
450 455 460

Ala Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly
465 470 475 480

His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe
485 490 495

Asp Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys
500 505 510

Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro
515 520 525

Ile Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser
530 535 540

Thr Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg
545 550 555 560

Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser
565 570 575

```
Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly
            580                 585                 590

Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val
        595                 600                 605

Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser
    610                 615                 620

Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His
625                 630                 635                 640

Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp
                645                 650                 655

Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr
            660                 665                 670

Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
        675                 680                 685

Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
    690                 695                 700

Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr
705                 710                 715                 720

Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala
                725                 730                 735

Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
            740                 745                 750

Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
        755                 760                 765

Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val
    770                 775                 780

His Asn Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val
785                 790                 795                 800

Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val
                805                 810                 815

Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
            820                 825                 830

Glu
```

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
- (A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

```
TGGCTATAGR CCAGGGCCAC                                                   20
```

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2505 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"
(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..2499
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

```
ATG AAT TCG GGG ATG CTG CCC CTC TTT GAG CCC AAG GGC CGG GTC CTC      48
Met Asn Ser Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu
  1               5                  10                  15

CTG GTG GAC GGC CAC CAC CTG GCC TAC CGC ACC TTC CAC GCC CTG AAG      96
Leu Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys
             20                  25                  30

GGC CTC ACC ACC AGC CGG GGG GAG CCG GTG CAG GCG GTC TAC GGC TTC     144
Gly Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe
         35                  40                  45

GCC AAG AGC CTC CTC AAG GCC CTC AAG GAG GAC GGG GAC GCG GTG ATC     192
Ala Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile
     50                  55                  60

GTG GTC TTT GAC GCC AAG GCC CCC TCC TTC CGC CAC GAG GCC TAC GGG     240
Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly
 65                  70                  75                  80

GGG TAC AAG GCG GGC CGG GCC CCC ACG CCG GAG GAC TTT CCC CGG CAA     288
Gly Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
                 85                  90                  95

CTC GCC CTC ATC AAG GAG CTG GTG GAC CTC CTG GGG CTG GCG CGC CTC     336
Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu
            100                 105                 110

GAG GTC CCG GGC TAC GAG GCG GAC GAC GTC CTG GCC AGC CTG GCC AAG     384
Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys
        115                 120                 125

AAG GCG GAA AAG GAG GGC TAC GAG GTC CGC ATC CTC ACC GCC GAC AAA     432
Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys
    130                 135                 140

GAC CTT TAC CAG CTC CTT TCC GAC CGC ATC CAC GTC CTC CAC CCC GAG     480
Asp Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu
145                 150                 155                 160

GGG TAC CTC ATC ACC CCG GCC TGG CTT TGG GAA AAG TAC GGC CTG AGG     528
Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg
                165                 170                 175

CCC GAC CAG TGG GCC GAC TAC CGG GCC CTG ACC GGG GAC GAG TCC GAC     576
Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp
            180                 185                 190

AAC CTT CCC GGG GTC AAG GGC ATC GGG GAG AAG ACG GCG AGG AAG CTT     624
Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu
        195                 200                 205
```

```
CTG GAG GAG TGG GGG AGC CTG GAA GCC CTC CTC AAG AAC CTG GAC CGG      672
Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg
    210                 215                 220

CTG AAG CCC GCC ATC CGG GAG AAG ATC CTG GCC CAC ATG GAC GAT CTG      720
Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu
225                 230                 235                 240

AAG CTC TCC TGG GAC CTG GCC AAG GTG CGC ACC GAC CTG CCC CTG GAG      768
Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu
                245                 250                 255

GTG GAC TTC GCC AAA AGG CGG GAG CCC GAC CGG GAG AGG CTT AGG GCC      816
Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala
            260                 265                 270

TTT CTG GAG AGG CTT GAG TTT GGC AGC CTC CTC CAC GAG TTC GGC CTT      864
Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu
        275                 280                 285

CTG GAA AGC CCC AAG GCC CTG GAG GAG GCC CCC TGG CCC CCG CCG GAA      912
Leu Glu Ser Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu
    290                 295                 300

GGG GCC TTC GTG GGC TTT GTG CTT TCC CGC AAG GAG CCC ATG TGG GCC      960
Gly Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala
305                 310                 315                 320

GAT CTT CTG GCC CTG GCC GCC GCC AGG GGG GGC CGG GTC CAC CGG GCC     1008
Asp Leu Leu Ala Leu Ala Ala Ala Arg Gly Gly Arg Val His Arg Ala
                325                 330                 335

CCC GAG CCT TAT AAA GCC CTC AGG GAC CTG AAG GAG GCG CGG GGG CTT     1056
Pro Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu
            340                 345                 350

CTC GCC AAA GAC CTG AGC GTT CTG GCC CTG AGG GAA GGC CTT GGC CTC     1104
Leu Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu
        355                 360                 365

CCG CCC GGC GAC GAC CCC ATG CTC CTC GCC TAC CTC CTG GAC CCT TCC     1152
Pro Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser
    370                 375                 380

AAC ACC ACC CCC GAG GGG GTG GCC CGG CGC TAC GGC GGG GAG TGG ACG     1200
Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr
385                 390                 395                 400

GAG GAG GCG GGG GAG CGG GCC GCC CTT TCC GAG AGG CTC TTC GCC AAC     1248
Glu Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn
                405                 410                 415

CTG TGG GGG AGG CTT GAG GGG GAG GAG AGG CTC CTT TGG CTT TAC CGG     1296
Leu Trp Gly Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Arg
            420                 425                 430

GAG GTG GAG AGG CCC CTT TCC GCT GTC CTG GCC CAC ATG GAG GCC ACG     1344
Glu Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr
        435                 440                 445

GGG GTG CGC CTG GAC GTG GCC TAT CTC AGG GCC TTG TCC CTG GAG GTG     1392
Gly Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val
    450                 455                 460

GCC GGG GAG ATC GCC CGC CTC GAG GCC GAG GTC TTC CGC CTG GCC GGC     1440
Ala Gly Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly
465                 470                 475                 480
```

CAC CCC TTC AAC CTC AAC TCC CGG GAC CAG CTG GAA AGG GTC CTC TTT 1488
His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe
485 490 495

GAC GAG CTA GGG CTT CCC GCC ATC GGC AAG ACG GAG AAG ACC GGC AAG 1536
Asp Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys
500 505 510

CGC TCC ACC AGC GCC GCC GTC CTG GAG GCC CTC CGC GAG GCC CAC CCC 1584
Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro
515 520 525

ATC GTG GAG AAG ATC CTG CAG TAC CGG GAG CTC ACC AAG CTG AAG AGC 1632
Ile Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser
530 535 540

ACC TAC ATT GAC CCC TTG CCG GAC CTC ATC CAC CCC AGG ACG GGC CGC 1680
Thr Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg
545 550 555 560

CTC CAC ACC CGC TTC AAC CAG ACG GCC ACG GCC ACG GGC AGG CTA AGT 1728
Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser
565 570 575

AGC TCC GAT CCC AAC CTC CAG AAC ATC CCC GTC CGC ACC CCG CTT GGG 1776
Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly
580 585 590

CAG AGG ATC CGC CGG GCC TTC ATC GCC GAG GAG GGG TGG CTA TTG GTG 1824
Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val
595 600 605

GCC CTG GCC TAT AGC CAG ATA GAG CTC AGG GTG CTG GCC CAC CTC TCC 1872
Ala Leu Ala Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser
610 615 620

GGC GAC GAG AAC CTG ATC CGG GTC TTC CAG GAG GGG CGG GAC ATC CAC 1920
Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His
625 630 635 640

ACG GAG ACC GCC AGC TGG ATG TTC GGC GTC CCC CGG GAG GCC GTG GAC 1968
Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp
645 650 655

CCC CTG ATG CGC CGG GCG GCC AAG ACC ATC AAC TTC GGG GTC CTC TAC 2016
Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr
660 665 670

GGC ATG TCG GCC CAC CGC CTC TCC CAG GAG CTA GCC ATC CCT TAC GAG 2064
Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
675 680 685

GAG GCC CAG GCC TTC ATT GAG CGC TAC TTT CAG AGC TTC CCC AAG GTG 2112
Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
690 695 700

CGG GCC TGG ATT GAG AAG ACC CTG GAG GAG GGC AGG AGG CGG GGG TAC 2160
Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr
705 710 715 720

GTG GAG ACC CTC TTC GGC CGC CGC CGC TAC GTG CCA GAC CTA GAG GCC 2208
Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala
725 730 735

CGG GTG AAG AGC GTG CGG GAG GCG GCC GAG CGC ATG GCC TTC AAC ATG 2256
Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
740 745 750

```
CCC GTC CAG GGC ACC GCC GCC GAC CTC ATG AAG CTG GCT ATG GTG AAG      2304
Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
        755                 760                 765

CTC TTC CCC AGG CTG GAG GAA ATG GGG GCC AGG ATG CTC CTT CAG GTC      2352
Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val
    770                 775                 780

CAC GAC GAG CTG GTC CTC GAG GCC CCA AAA GAG AGG GCG GAG GCC GTG      2400
His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val
785                 790                 795                 800

GCC CGG CTG GCC AAG GAG GTC ATG GAG GGG GTG TAT CCC CTG GCC GTG      2448
Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val
                805                 810                 815

CCC CTG GAG GTG GAG GTG GGG ATA GGG GAG GAC TGG CTC TCC GCC AAG      2496
Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
            820                 825                 830

GAG TGATAG                                                            2505
Glu
```

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 833 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

```
Met Asn Ser Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu
  1               5                  10                  15

Leu Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys
             20                  25                  30

Gly Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe
         35                  40                  45

Ala Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile
     50                  55                  60

Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly
 65                  70                  75                  80

Gly Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
                 85                  90                  95

Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu
            100                 105                 110

Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys
        115                 120                 125

Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys
    130                 135                 140

Asp Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu
145                 150                 155                 160

Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg
                165                 170                 175
```

```
Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp
            180                 185                 190

Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu
        195                 200                 205

Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg
    210                 215                 220

Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu
225                 230                 235                     240

Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu
                245                 250                 255

Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala
            260                 265                 270

Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu
        275                 280                 285

Leu Glu Ser Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu
    290                 295                 300

Gly Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala
305                 310                 315                     320

Asp Leu Leu Ala Leu Ala Ala Ala Arg Gly Gly Arg Val His Arg Ala
                325                 330                 335

Pro Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu
            340                 345                 350

Leu Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu
        355                 360                 365

Pro Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser
    370                 375                 380

Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr
385                 390                 395                     400

Glu Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn
                405                 410                 415

Leu Trp Gly Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Arg
            420                 425                 430

Glu Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr
        435                 440                 445

Gly Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val
    450                 455                 460

Ala Gly Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly
465                 470                 475                     480

His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe
                485                 490                 495

Asp Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys
            500                 505                 510

Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro
        515                 520                 525
```

```
Ile Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser
    530                 535                 540

Thr Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg
545                 550                 555                 560

Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser
                565                 570                 575

Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly
            580                 585                 590

Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val
        595                 600                 605

Ala Leu Ala Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser
    610                 615                 620

Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His
625                 630                 635                 640

Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp
                645                 650                 655

Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr
            660                 665                 670

Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
        675                 680                 685

Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
    690                 695                 700

Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr
705                 710                 715                 720

Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala
                725                 730                 735

Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
            740                 745                 750

Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
        755                 760                 765

Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val
    770                 775                 780

His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val
785                 790                 795                 800

Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val
                805                 810                 815

Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
            820                 825                 830

Glu
```

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2505 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

```
     (ii) MOLECULE TYPE: other nucleic acid
          (A) DESCRIPTION: /desc = "DNA"

     (ix) FEATURE:
          (A) NAME/KEY: CDS
          (B) LOCATION: 1..2499

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

ATG AAT TCG GGG ATG CTG CCC CTC TTT GAG CCC AAG GGC CGG GTC CTC        48
Met Asn Ser Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu
  1               5                  10                  15

CTG GTG GAC GGC CAC CAC CTG GCC TAC CGC ACC TTC CAC GCC CTG AAG        96
Leu Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys
             20                  25                  30

GGC CTC ACC ACC AGC CGG GGG GAG CCG GTG CAG GCG GTC TAC GGC TTC       144
Gly Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe
         35                  40                  45

GCC AAG AGC CTC CTC AAG GCC CTC AAG GAG GAC GGG GAC GCG GTG ATC       192
Ala Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile
     50                  55                  60

GTG GTC TTT GAC GCC AAG GCC CCC TCC TTC CGC CAC GAG GCC TAC GGG       240
Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly
 65                  70                  75                  80

GGG TAC AAG GCG GGC CGG GCC CCC ACG CCG GAG GAC TTT CCC CGG CAA       288
Gly Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
                 85                  90                  95

CTC GCC CTC ATC AAG GAG CTG GTG GAC CTC CTG GGG CTG GCG CGC CTC       336
Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu
            100                 105                 110

GAG GTC CCG GGC TAC GAG GCG GAC GAC GTC CTG GCC AGC CTG GCC AAG       384
Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys
        115                 120                 125

AAG GCG GAA AAG GAG GGC TAC GAG GTC CGC ATC CTC ACC GCC GAC AAA       432
Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys
    130                 135                 140

GAC CTT TAC CAG CTC CTT TCC GAC CGC ATC CAC GTC CTC CAC CCC GAG       480
Asp Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu
145                 150                 155                 160

GGG TAC CTC ATC ACC CCG GCC TGG CTT TGG GAA AAG TAC GGC CTG AGG       528
Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg
                165                 170                 175

CCC GAC CAG TGG GCC GAC TAC CGG GCC CTG ACC GGG GAC GAG TCC GAC       576
Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp
            180                 185                 190

AAC CTT CCC GGG GTC AAG GGC ATC GGG GAG AAG ACG GCG AGG AAG CTT       624
Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu
        195                 200                 205

CTG GAG GAG TGG GGG AGC CTG GAA GCC CTC CTC AAG AAC CTG GAC CGG       672
Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg
    210                 215                 220
```

```
CTG AAG CCC GCC ATC CGG GAG AAG ATC CTG GCC CAC ATG GAC GAT CTG        720
Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu
225                 230                 235                 240

AAG CTC TCC TGG GAC CTG GCC AAG GTG CGC ACC GAC CTG CCC CTG GAG        768
Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu
                245                 250                 255

GTG GAC TTC GCC AAA AGG CGG GAG CCC GAC CGG GAG AGG CTT AGG GCC        816
Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala
            260                 265                 270

TTT CTG GAG AGG CTT GAG TTT GGC AGC CTC CTC CAC GAG TTC GGC CTT        864
Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu
        275                 280                 285

CTG GAA AGC CCC AAG GCC CTG GAG GAG GCC CCC TGG CCC CCG CCG GAA        912
Leu Glu Ser Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu
    290                 295                 300

GGG GCC TTC GTG GGC TTT GTG CTT TCC CGC AAG GAG CCC ATG TGG GCC        960
Gly Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala
305                 310                 315                 320

GAT CTT CTG GCC CTG GCC GCC GCC AGG GGG GGC CGG GTC CAC CGG GCC       1008
Asp Leu Leu Ala Leu Ala Ala Ala Arg Gly Gly Arg Val His Arg Ala
                325                 330                 335

CCC GAG CCT TAT AAA GCC CTC AGG GAC CTG AAG GAG GCG CGG GGG CTT       1056
Pro Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu
            340                 345                 350

CTC GCC AAA GAC CTG AGC GTT CTG GCC CTG AGG GAA GGC CTT GGC CTC       1104
Leu Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu
        355                 360                 365

CCG CCC GGC GAC GAC CCC ATG CTC CTC GCC TAC CTC CTG GAC CCT TCC       1152
Pro Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser
    370                 375                 380

AAC ACC ACC CCC GAG GGG GTG GCC CGG CGC TAC GGC GGG GAG TGG ACG       1200
Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr
385                 390                 395                 400

GAG GAG GCG GGG GAG CGG GCC GCC CTT TCC GAG AGG CTC TTC GCC AAC       1248
Glu Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn
                405                 410                 415

CTG TGG GGG AGG CTT GAG GGG GAG GAG AGG CTC CTT TGG CTT TAC CGG       1296
Leu Trp Gly Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Arg
            420                 425                 430

GAG GTG GAG AGG CCC CTT TCC GCT GTC CTG GCC CAC ATG GAG GCC ACG       1344
Glu Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr
        435                 440                 445

GGG GTG CGC CTG GAC GTG GCC TAT CTC AGG GCC TTG TCC CTG GAG GTG       1392
Gly Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val
    450                 455                 460

GCC GGG GAG ATC GCC CGC CTC GAG GCC GAG GTC TTC CGC CTG GCC GGC       1440
Ala Gly Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly
465                 470                 475                 480

CAC CCC TTC AAC CTC AAC TCC CGG GAC CAG CTG GAA AGG GTC CTC TTT       1488
His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe
                485                 490                 495
```

```
GAC GAG CTA GGG CTT CCC GCC ATC GGC AAG ACG GAG AAG ACC GGC AAG      1536
Asp Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys
            500                 505                 510

CGC TCC ACC AGC GCC GCC GTC CTG GAG GCC CTC CGC GAG GCC CAC CCC      1584
Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro
        515                 520                 525

ATC GTG GAG AAG ATC CTG CAG TAC CGG GAG CTC ACC AAG CTG AAG AGC      1632
Ile Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser
    530                 535                 540

ACC TAC ATT GAC CCC TTG CCG GAC CTC ATC CAC CCC AGG ACG GGC CGC      1680
Thr Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg
545                 550                 555                 560

CTC CAC ACC CGC TTC AAC CAG ACG GCC ACG GCC ACG GGC AGG CTA AGT      1728
Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser
                565                 570                 575

AGC TCC GAT CCC AAC CTC CAG AAC ATC CCC GTC CGC ACC CCG CTT GGG      1776
Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly
            580                 585                 590

CAG AGG ATC CGC CGG GCC TTC ATC GCC GAG GAG GGG TGG CTA TTG GTG      1824
Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val
        595                 600                 605

GCC CTG GTC TAT AGC CAG ATA GAG CTC AGG GTG CTG GCC CAC CTC TCC      1872
Ala Leu Val Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser
    610                 615                 620

GGC GAC GAG AAC CTG ATC CGG GTC TTC CAG GAG GGG CGG GAC ATC CAC      1920
Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His
625                 630                 635                 640

ACG GAG ACC GCC AGC TGG ATG TTC GGC GTC CCC CGG GAG GCC GTG GAC      1968
Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp
                645                 650                 655

CCC CTG ATG CGC CGG GCG GCC AAG ACC ATC AAC TTC GGG GTC CTC TAC      2016
Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr
            660                 665                 670

GGC ATG TCG GCC CAC CGC CTC TCC CAG GAG CTA GCC ATC CCT TAC GAG      2064
Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
        675                 680                 685

GAG GCC CAG GCC TTC ATT GAG CGC TAC TTT CAG AGC TTC CCC AAG GTG      2112
Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
    690                 695                 700

CGG GCC TGG ATT GAG AAG ACC CTG GAG GAG GGC AGG AGG CGG GGG TAC      2160
Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr
705                 710                 715                 720

GTG GAG ACC CTC TTC GGC CGC CGC CGC TAC GTG CCA GAC CTA GAG GCC      2208
Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala
                725                 730                 735

CGG GTG AAG AGC GTG CGG GAG GCG GCC GAG CGC ATG GCC TTC AAC ATG      2256
Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
            740                 745                 750

CCC GTC CAG GGC ACC GCC GCC GAC CTC ATG AAG CTG GCT ATG GTG AAG      2304
Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
        755                 760                 765
```

```
CTC TTC CCC AGG CTG GAG GAA ATG GGG GCC AGG ATG CTC CTT CAG GTC      2352
Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val
    770                 775                 780

CAC GAC GAG CTG GTC CTC GAG GCC CCA AAA GAG AGG GCG GAG GCC GTG      2400
His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val
785                 790                 795                 800

GCC CGG CTG GCC AAG GAG GTC ATG GAG GGG GTG TAT CCC CTG GCC GTG      2448
Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val
                805                 810                 815

CCC CTG GAG GTG GAG GTG GGG ATA GGG GAG GAC TGG CTC TCC GCC AAG      2496
Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
            820                 825                 830

GAG TGATAG                                                           2505
Glu
```

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 833 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

```
Met Asn Ser Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu
  1               5                  10                  15

Leu Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys
             20                  25                  30

Gly Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe
         35                  40                  45

Ala Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile
     50                  55                  60

Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly
 65                  70                  75                  80

Gly Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
                 85                  90                  95

Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu
            100                 105                 110

Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys
        115                 120                 125

Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys
    130                 135                 140

Asp Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu
145                 150                 155                 160

Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg
                165                 170                 175

Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp
            180                 185                 190
```

Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu
195 200 205

Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg
210 215 220

Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu
225 230 235 240

Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu
245 250 255

Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala
260 265 270

Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu
275 280 285

Leu Glu Ser Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu
290 295 300

Gly Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala
305 310 315 320

Asp Leu Leu Ala Leu Ala Ala Ala Arg Gly Gly Arg Val His Arg Ala
325 330 335

Pro Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu
340 345 350

Leu Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu
355 360 365

Pro Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser
370 375 380

Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr
385 390 395 400

Glu Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn
405 410 415

Leu Trp Gly Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Arg
420 425 430

Glu Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr
435 440 445

Gly Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val
450 455 460

Ala Gly Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly
465 470 475 480

His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe
485 490 495

Asp Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys
500 505 510

Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro
515 520 525

Ile Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser
530 535 540

```
Thr Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg
545                 550                 555                 560

Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser
                565                 570                 575

Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly
            580                 585                 590

Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val
        595                 600                 605

Ala Leu Val Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser
    610                 615                 620

Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His
625                 630                 635                 640

Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp
                645                 650                 655

Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr
            660                 665                 670

Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
        675                 680                 685

Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
    690                 695                 700

Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr
705                 710                 715                 720

Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala
                725                 730                 735

Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
            740                 745                 750

Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
        755                 760                 765

Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val
    770                 775                 780

His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val
785                 790                 795                 800

Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val
                805                 810                 815

Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
            820                 825                 830

Glu
```

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

GGGATACCAT GGGAGTGCAG TTTGG 25

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

GGTAAATTTT TCTCGTCGAC ATCCCAC 27

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 981 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..978

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

```
ATG GGA GTG CAG TTT GGT GAT TTT ATT CCA AAA AAT ATT ATC TCC TTT       48
Met Gly Val Gln Phe Gly Asp Phe Ile Pro Lys Asn Ile Ile Ser Phe
  1               5                  10                  15

GAA GAT TTA AAA GGG AAA AAA GTA GCT ATT GAT GGA ATG AAT GCA TTA       96
Glu Asp Leu Lys Gly Lys Lys Val Ala Ile Asp Gly Met Asn Ala Leu
             20                  25                  30

TAT CAG TTT TTA ACA TCT ATA CGT TTG AGA GAT GGT TCT CCA TTG AGA      144
Tyr Gln Phe Leu Thr Ser Ile Arg Leu Arg Asp Gly Ser Pro Leu Arg
         35                  40                  45

AAT AGA AAA GGA GAG ATA ACC TCA GCA TAT AAC GGA GTT TTT TAT AAA      192
Asn Arg Lys Gly Glu Ile Thr Ser Ala Tyr Asn Gly Val Phe Tyr Lys
     50                  55                  60

ACC ATA CAT TTG TTA GAG AAT GAT ATA ACT CCA ATC TGG GTT TTT GAT      240
Thr Ile His Leu Leu Glu Asn Asp Ile Thr Pro Ile Trp Val Phe Asp
 65                  70                  75                  80

GGT GAG CCA CCA AAG TTA AAG GAG AAA ACA AGG AAA GTT AGG AGA GAG      288
Gly Glu Pro Pro Lys Leu Lys Glu Lys Thr Arg Lys Val Arg Arg Glu
                 85                  90                  95

ATG AAA GAG AAA GCT GAA CTT AAG ATG AAA GAG GCA ATT AAA AAG GAG      336
Met Lys Glu Lys Ala Glu Leu Lys Met Lys Glu Ala Ile Lys Lys Glu
            100                 105                 110

GAT TTT GAA GAA GCT GCT AAG TAT GCA AAG AGG GTT AGC TAT CTA ACT      384
Asp Phe Glu Glu Ala Ala Lys Tyr Ala Lys Arg Val Ser Tyr Leu Thr
        115                 120                 125
```

```
CCG AAA ATG GTT GAA AAC TGC AAA TAT TTG TTA AGT TTG ATG GGC ATT        432
Pro Lys Met Val Glu Asn Cys Lys Tyr Leu Leu Ser Leu Met Gly Ile
    130                 135                 140

CCG TAT GTT GAA GCT CCC TCT GAG GGA GAG GCA CAA GCA AGC TAT ATG        480
Pro Tyr Val Glu Ala Pro Ser Glu Gly Glu Ala Gln Ala Ser Tyr Met
145                 150                 155                 160

GCA AAG AAG GGA GAT GTT TGG GCA GTT GTA AGT CAA GAT TAT GAT GCC        528
Ala Lys Lys Gly Asp Val Trp Ala Val Val Ser Gln Asp Tyr Asp Ala
                165                 170                 175

TTG TTA TAT GGA GCT CCG AGA GTT GTT AGA AAT TTA ACA ACT ACA AAG        576
Leu Leu Tyr Gly Ala Pro Arg Val Val Arg Asn Leu Thr Thr Thr Lys
            180                 185                 190

GAG ATG CCA GAA CTT ATT GAA TTA AAT GAG GTT TTA GAG GAT TTA AGA        624
Glu Met Pro Glu Leu Ile Glu Leu Asn Glu Val Leu Glu Asp Leu Arg
        195                 200                 205

ATT TCT TTG GAT GAT TTG ATA GAT ATA GCC ATA TTT ATG GGA ACT GAC        672
Ile Ser Leu Asp Asp Leu Ile Asp Ile Ala Ile Phe Met Gly Thr Asp
    210                 215                 220

TAT AAT CCA GGA GGA GTT AAA GGA ATA GGA TTT AAA AGG GCT TAT GAA        720
Tyr Asn Pro Gly Gly Val Lys Gly Ile Gly Phe Lys Arg Ala Tyr Glu
225                 230                 235                 240

TTG GTT AGA AGT GGT GTA GCT AAG GAT GTT TTG AAA AAA GAG GTT GAA        768
Leu Val Arg Ser Gly Val Ala Lys Asp Val Leu Lys Lys Glu Val Glu
                245                 250                 255

TAC TAC GAT GAG ATT AAG AGG ATA TTT AAA GAG CCA AAG GTT ACC GAT        816
Tyr Tyr Asp Glu Ile Lys Arg Ile Phe Lys Glu Pro Lys Val Thr Asp
            260                 265                 270

AAC TAT TCA TTA AGC CTA AAA TTG CCA GAT AAA GAG GGA ATT ATA AAA        864
Asn Tyr Ser Leu Ser Leu Lys Leu Pro Asp Lys Glu Gly Ile Ile Lys
        275                 280                 285

TTC TTA GTT GAT GAA AAT GAC TTT AAT TAT GAT AGG GTT AAA AAG CAT        912
Phe Leu Val Asp Glu Asn Asp Phe Asn Tyr Asp Arg Val Lys Lys His
    290                 295                 300

GTT GAT AAA CTC TAT AAC TTA ATT GCA AAC AAA ACT AAG CAA AAA ACA        960
Val Asp Lys Leu Tyr Asn Leu Ile Ala Asn Lys Thr Lys Gln Lys Thr
305                 310                 315                 320

TTA GAT GCA TGG TTT AAA TAA                                            981
Leu Asp Ala Trp Phe Lys
                325
```

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 326 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

```
Met Gly Val Gln Phe Gly Asp Phe Ile Pro Lys Asn Ile Ile Ser Phe
 1               5                  10                  15

Glu Asp Leu Lys Gly Lys Lys Val Ala Ile Asp Gly Met Asn Ala Leu
            20                  25                  30
```

Tyr Gln Phe Leu Thr Ser Ile Arg Leu Arg Asp Gly Ser Pro Leu Arg
35 40 45

Asn Arg Lys Gly Glu Ile Thr Ser Ala Tyr Asn Gly Val Phe Tyr Lys
50 55 60

Thr Ile His Leu Leu Glu Asn Asp Ile Thr Pro Ile Trp Val Phe Asp
65 70 75 80

Gly Glu Pro Pro Lys Leu Lys Glu Lys Thr Arg Lys Val Arg Arg Glu
85 90 95

Met Lys Glu Lys Ala Glu Leu Lys Met Lys Glu Ala Ile Lys Lys Glu
100 105 110

Asp Phe Glu Glu Ala Ala Lys Tyr Ala Lys Arg Val Ser Tyr Leu Thr
115 120 125

Pro Lys Met Val Glu Asn Cys Lys Tyr Leu Leu Ser Leu Met Gly Ile
130 135 140

Pro Tyr Val Glu Ala Pro Ser Glu Gly Glu Ala Gln Ala Ser Tyr Met
145 150 155 160

Ala Lys Lys Gly Asp Val Trp Ala Val Val Ser Gln Asp Tyr Asp Ala
165 170 175

Leu Leu Tyr Gly Ala Pro Arg Val Val Arg Asn Leu Thr Thr Thr Lys
180 185 190

Glu Met Pro Glu Leu Ile Glu Leu Asn Glu Val Leu Glu Asp Leu Arg
195 200 205

Ile Ser Leu Asp Asp Leu Ile Asp Ile Ala Ile Phe Met Gly Thr Asp
210 215 220

Tyr Asn Pro Gly Gly Val Lys Gly Ile Gly Phe Lys Arg Ala Tyr Glu
225 230 235 240

Leu Val Arg Ser Gly Val Ala Lys Asp Val Leu Lys Lys Glu Val Glu
245 250 255

Tyr Tyr Asp Glu Ile Lys Arg Ile Phe Lys Glu Pro Lys Val Thr Asp
260 265 270

Asn Tyr Ser Leu Ser Leu Lys Leu Pro Asp Lys Glu Gly Ile Ile Lys
275 280 285

Phe Leu Val Asp Glu Asn Asp Phe Asn Tyr Asp Arg Val Lys Lys His
290 295 300

Val Asp Lys Leu Tyr Asn Leu Ile Ala Asn Lys Thr Lys Gln Lys Thr
305 310 315 320

Leu Asp Ala Trp Phe Lys
325

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

GAGGTGATAC CATGGGTGTC C 21

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

GAAACTCTGC AGCGCGTCAG 20

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1023 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)
(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1020
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

```
ATG GGT GTC CCA ATT GGT GAG ATT ATA CCA AGA AAA GAA ATT GAG TTA       48
Met Gly Val Pro Ile Gly Glu Ile Ile Pro Arg Lys Glu Ile Glu Leu
  1               5                  10                  15

GAA AAC CTA TAC GGG AAA AAA ATC GCA ATC GAC GCT CTT AAT GCA ATC       96
Glu Asn Leu Tyr Gly Lys Lys Ile Ala Ile Asp Ala Leu Asn Ala Ile
             20                  25                  30

TAC CAA TTT TTG TCC ACA ATA AGA CAG AAA GAT GGA ACT CCA CTT ATG      144
Tyr Gln Phe Leu Ser Thr Ile Arg Gln Lys Asp Gly Thr Pro Leu Met
         35                  40                  45

GAT TCA AAG GGT AGA ATA ACC TCC CAC CTA AGC GGG CTC TTT TAC AGG      192
Asp Ser Lys Gly Arg Ile Thr Ser His Leu Ser Gly Leu Phe Tyr Arg
     50                  55                  60

ACA ATA AAC CTA ATG GAG GCT GGA ATA AAA CCT GTG TAT GTT TTT GAT      240
Thr Ile Asn Leu Met Glu Ala Gly Ile Lys Pro Val Tyr Val Phe Asp
 65                  70                  75                  80

GGA GAA CCT CCA GAA TTC AAA AAG AAA GAG CTC GAA AAA AGA AGA GAA      288
Gly Glu Pro Pro Glu Phe Lys Lys Lys Glu Leu Glu Lys Arg Arg Glu
                 85                  90                  95

GCG AGA GAG GAA GCT GAA GAA AAG TGG AGA GAA GCA CTT GAA AAA GGA      336
Ala Arg Glu Glu Ala Glu Glu Lys Trp Arg Glu Ala Leu Glu Lys Gly
            100                 105                 110

GAG ATA GAG GAA GCA AGA AAA TAT GCC CAA AGA GCA ACC AGG GTA AAT      384
Glu Ile Glu Glu Ala Arg Lys Tyr Ala Gln Arg Ala Thr Arg Val Asn
        115                 120                 125
```

```
GAA ATG CTC ATC GAG GAT GCA AAA AAA CTC TTA GAG CTT ATG GGA ATT         432
Glu Met Leu Ile Glu Asp Ala Lys Lys Leu Leu Glu Leu Met Gly Ile
    130                 135                     140

CCT ATA GTT CAA GCA CCT AGC GAG GGA GAG GCC CAA GCT GCA TAT ATG         480
Pro Ile Val Gln Ala Pro Ser Glu Gly Glu Ala Gln Ala Ala Tyr Met
145                 150                     155                 160

GCC GCA AAG GGG AGC GTG TAT GCA TCG GCT AGT CAA GAT TAC GAT TCC         528
Ala Ala Lys Gly Ser Val Tyr Ala Ser Ala Ser Gln Asp Tyr Asp Ser
                165                     170                 175

CTA CTT TTT GGA GCT CCA AGA CTT GTT AGA AAC TTA ACA ATA ACA GGA         576
Leu Leu Phe Gly Ala Pro Arg Leu Val Arg Asn Leu Thr Ile Thr Gly
            180                 185                     190

AAA AGA AAG TTG CCT GGG AAA AAT GTC TAC GTC GAG ATA AAG CCC GAG         624
Lys Arg Lys Leu Pro Gly Lys Asn Val Tyr Val Glu Ile Lys Pro Glu
        195                 200                     205

TTG ATA ATT TTG GAG GAA GTA CTC AAG GAA TTA AAG CTA ACA AGA GAA         672
Leu Ile Ile Leu Glu Glu Val Leu Lys Glu Leu Lys Leu Thr Arg Glu
    210                 215                     220

AAG CTC ATT GAA CTA GCA ATC CTC GTT GGA ACA GAC TAC AAC CCA GGA         720
Lys Leu Ile Glu Leu Ala Ile Leu Val Gly Thr Asp Tyr Asn Pro Gly
225                 230                     235                 240

GGA ATA AAG GGC ATA GGC CTT AAA AAA GCT TTA GAG ATT GTT AGA CAC         768
Gly Ile Lys Gly Ile Gly Leu Lys Lys Ala Leu Glu Ile Val Arg His
                245                     250                 255

TCA AAA GAT CCG CTA GCA AAG TTC CAA AAG CAA AGC GAT GTG GAT TTA         816
Ser Lys Asp Pro Leu Ala Lys Phe Gln Lys Gln Ser Asp Val Asp Leu
            260                 265                     270

TAT GCA ATA AAA GAG TTC TTC CTA AAC CCA CCA GTC ACA GAT AAC TAC         864
Tyr Ala Ile Lys Glu Phe Phe Leu Asn Pro Pro Val Thr Asp Asn Tyr
        275                 280                     285

AAT TTA GTG TGG AGA GAT CCC GAC GAA GAG GGA ATA CTA AAG TTC TTA         912
Asn Leu Val Trp Arg Asp Pro Asp Glu Glu Gly Ile Leu Lys Phe Leu
    290                 295                     300

TGT GAC GAG CAT GAC TTT AGT GAG GAA AGA GTA AAG AAT GGA TTA GAG         960
Cys Asp Glu His Asp Phe Ser Glu Glu Arg Val Lys Asn Gly Leu Glu
305                 310                     315                 320

AGG CTT AAG AAG GCA ATC AAA AGT GGA AAA CAA TCA ACC CTT GAA AGT        1008
Arg Leu Lys Lys Ala Ile Lys Ser Gly Lys Gln Ser Thr Leu Glu Ser
                325                     330                 335

TGG TTC AAG AGA TAA                                                    1023
Trp Phe Lys Arg
            340
```

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 340 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

```
Met Gly Val Pro Ile Gly Glu Ile Ile Pro Arg Lys Glu Ile Glu Leu
 1               5                  10                  15

Glu Asn Leu Tyr Gly Lys Lys Ile Ala Ile Asp Ala Leu Asn Ala Ile
            20                  25                  30

Tyr Gln Phe Leu Ser Thr Ile Arg Gln Lys Asp Gly Thr Pro Leu Met
        35                  40                  45

Asp Ser Lys Gly Arg Ile Thr Ser His Leu Ser Gly Leu Phe Tyr Arg
    50                  55                  60

Thr Ile Asn Leu Met Glu Ala Gly Ile Lys Pro Val Tyr Val Phe Asp
65                  70                  75                  80

Gly Glu Pro Pro Glu Phe Lys Lys Lys Glu Leu Glu Lys Arg Arg Glu
                85                  90                  95

Ala Arg Glu Glu Ala Glu Glu Lys Trp Arg Glu Ala Leu Glu Lys Gly
            100                 105                 110

Glu Ile Glu Glu Ala Arg Lys Tyr Ala Gln Arg Ala Thr Arg Val Asn
        115                 120                 125

Glu Met Leu Ile Glu Asp Ala Lys Lys Leu Leu Glu Leu Met Gly Ile
    130                 135                 140

Pro Ile Val Gln Ala Pro Ser Glu Gly Glu Ala Gln Ala Ala Tyr Met
145                 150                 155                 160

Ala Ala Lys Gly Ser Val Tyr Ala Ser Ala Ser Gln Asp Tyr Asp Ser
                165                 170                 175

Leu Leu Phe Gly Ala Pro Arg Leu Val Arg Asn Leu Thr Ile Thr Gly
            180                 185                 190

Lys Arg Lys Leu Pro Gly Lys Asn Val Tyr Val Glu Ile Lys Pro Glu
        195                 200                 205

Leu Ile Ile Leu Glu Glu Val Leu Lys Glu Leu Lys Leu Thr Arg Glu
    210                 215                 220

Lys Leu Ile Glu Leu Ala Ile Leu Val Gly Thr Asp Tyr Asn Pro Gly
225                 230                 235                 240

Gly Ile Lys Gly Ile Gly Leu Lys Lys Ala Leu Glu Ile Val Arg His
                245                 250                 255

Ser Lys Asp Pro Leu Ala Lys Phe Gln Lys Gln Ser Asp Val Asp Leu
            260                 265                 270

Tyr Ala Ile Lys Glu Phe Phe Leu Asn Pro Pro Val Thr Asp Asn Tyr
        275                 280                 285

Asn Leu Val Trp Arg Asp Pro Asp Glu Glu Gly Ile Leu Lys Phe Leu
    290                 295                 300

Cys Asp Glu His Asp Phe Ser Glu Glu Arg Val Lys Asn Gly Leu Glu
305                 310                 315                 320

Arg Leu Lys Lys Ala Ile Lys Ser Gly Lys Gln Ser Thr Leu Glu Ser
                325                 330                 335

Trp Phe Lys Arg
            340
```

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

GATACCATGG GTGTCCCAAT TGGTG 25

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

TCGACGTCGA CTTATCTCTT GAACCAACTT TCAAGGG 37

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

AGCGAGGGAG AGGCCCAAGC 20

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

GCCTATGCCC TTTATTCCTC C 21

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

TGGTCGCTGT CTCGCTGAAA GCGAGACAGC GTG 33

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

TGCTCTCTGG TCGCTGTCTG AAAGACAGCG 30

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"
(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a spacer containing a fluorescein label."
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

NAGAAAGGAA GGGAAGAAAG CGAAAGG 27

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"
(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(27, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a spacer bearing a Cy3 dye."

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 28
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "The residue at this position is a dideoxycytidine."
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

AGAAAGGAAG GGAAGAAAGC GAAAGGNC 28

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"
(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 24
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "The residue at this position is a dideoxycytidine."
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

GCCGGCGAAC GTGGCGAGAA AGGC 24

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a spacer containing a fluorescein label."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(28, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a spacer bearing a Cy3 dye."

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 29
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "The residue at this position is a dideoxycytidine."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

NAGAAAGGAA GGGAAGAAAG CGAAAGGNC 29

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

AAAATTCCTT TCTCTTTGCC CTTTGCTTCC 30

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

GGAAAGCCGG CGAACGTGGC GAGAAA 26

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

GGAAAGCCGG CGAACGTGGC GAGA 24

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a spacer bearing a Cy3 dye."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(28, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a spacer bearing a biotin group."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

NAGAAAGGAA GGGAAGAAAG CGAAAGGNT 29

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a spacer bearing a Cy3 dye."

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 2..3
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "The residues at these positions have an amino group added."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(24, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a spacer containing a fluorescein label."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

NTTCCAGAGC CTAATTTGCC AGTNA 25

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "The residue at this position has a 5' TET-label."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(23, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a spacer containing a fluorescein label."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

TTCCAGAGCC TAATTTGCCA GTNA 24

(2) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

CTTACCAACG CTAACGAGCG TCTTG 25

(2) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

CCCGTCTCGC TGGTGAAAAG AAAAACCACC CTGGCGCCCA ATA 43

(2) INFORMATION FOR SEQ ID NO:98:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

TATTGGGCGC CATGGTGGTT TTT 23

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(10, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 5-nitroindole."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(16, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 5-nitroindole."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

TATTGGGCGN CAGGGNGGTT TTT 23

(2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(10, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 5-nitroindole."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(16, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 5-nitroindole."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

TATTGGGCGN CATGGNGGTT TTT 23

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(16, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3-nitropyrrole."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

TATTGGGCGC CAGGGNGGTT TTT 23

(2) INFORMATION FOR SEQ ID NO:102:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(16, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3-nitropyrrole group."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

TATTGGGCGC CATGGNGGTT TTT 23

(2) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxycytosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(2, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxythymidine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(3, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxyguanosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(4..5, "")
(D) OTHER INFORMATION: /note= "The residues at these positions are a 2'deoxyadenosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(6, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxythymidine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(7, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxyadenosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(8, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxythymidine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(9..10, "")
(D) OTHER INFORMATION: /note= "The residues at these positions are a 2'deoxyadenosine 5'-O-(1-Thiomonophosphate)."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

CTGAATATAA ACTTGTGGTA GTTGGAGCTG GTGCCGTAGG CAAGAGTGCC TTGACG 56

(2) INFORMATION FOR SEQ ID NO:104:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxycytosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(2, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxythymidine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(3, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxyguanosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(4..5, "")
(D) OTHER INFORMATION: /note= "The residues at these positions are a 2'deoxyadenosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(6, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxythymidine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(7, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxyadenosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(8, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxythymidine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(9..10, "")
(D) OTHER INFORMATION: /note= "The residues at these positions are a 2'deoxyadenosine 5'-O-(1-Thiomonophosphate)."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

CTGAATATAA ACTTGTGGTA GTTGGAGCTG GTGACGTAGG CAAGAGTGCC TTGACG 56

(2) INFORMATION FOR SEQ ID NO:105:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxyguanosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(2, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxycytosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(3, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxythymidine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(4, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxycytosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(5..6, "")
(D) OTHER INFORMATION: /note= "The residues at these positions are 2'deoxyadenosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(7..8, "")
(D) OTHER INFORMATION: /note= "The residues at these positions are 2'deoxyguanosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(9, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxycytosine 5'-O-(1-Thiomonophosphate)."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

GCTCAAGGCA CTCTTGCCTA CGA 23

(2) INFORMATION FOR SEQ ID NO:106:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a spacer bearing a Cy3 amidite label."

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 2..3
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "The residues at these positions have an amino group added."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

NTTCACCAG 9

(2) INFORMATION FOR SEQ ID NO:107:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxycytosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(2, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxythymidine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(3..4, "")
(D) OTHER INFORMATION: /note= "The residues at these positions are a 2'deoxycytosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(5..6, "")
(D) OTHER INFORMATION: /note= "The residues at these positions are a 2'deoxyadenosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(7, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxycytosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(8, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxythymidine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(9, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxyadenosine 5'-O-(1-Thiomonophosphate)."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

CTCCAACTAC CACAAGTTTA TATTCAG 27

(2) INFORMATION FOR SEQ ID NO:108:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

CGAGAGACCA CGCT 14

(2) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(14, "")
(D) OTHER INFORMATION: /note= "The residue at this positions contain an abasic ribose."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

CGAGAGACCA CGCT 14

(2) INFORMATION FOR SEQ ID NO:110:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"
(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(14, "")
(D) OTHER INFORMATION: /note= "The residue at this position contains an abasic ribose with a 3' phosphate group."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

CGAGAGACCA CGCT 14

(2) INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(15, "")
(D) OTHER INFORMATION: /note= "The residue at this position contains a 3' phosphate group."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

CGAGAGACCA CGCTG 15

(2) INFORMATION FOR SEQ ID NO:112:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxyguanosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(2, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxythymidine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(3..4, "")
(D) OTHER INFORMATION: /note= "The residues at these positions are a 2'deoxyadenosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(5, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxythymidine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(6, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxycytosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(7..8, "")
(D) OTHER INFORMATION: /note= "The residues at these positions are a 2'deoxythymidine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(9, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxyadenosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(10, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxycytosine 5'-O-(1-Thiomonophosphate)."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

GTAATCTTAC CAACGCTAAC GAGCGTCTTG 30

(2) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1..2, "")
(D) OTHER INFORMATION: /note= "The residues at these positions are a 2'deoxycytosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(3, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxythymidine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(4..5, "")
(D) OTHER INFORMATION: /note= "The residues at these positions are a 2'deoxyadenosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(6..8, "")
(D) OTHER INFORMATION: /note= "The residues at these positions are a 2'deoxythymidine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(9, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxyguanosine 5'-O-(1-Thiomonophosphate)."

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(10, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 2'deoxycytosine 5'-O-(1-Thiomonophosphate)."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

CCTAATTTGC CAGTTACAAA ATAAACAGCC C 31

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a spacer bearing a Cy3 dye."

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 2..3
(D) OTHER INFORMATION: /mod_base= OTHER /note= "The residues at these positions have an amino group added."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

NTTCCAGAG 9

(2) INFORMATION FOR SEQ ID NO:115:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

TTTTCCAGAG CCTAATGAAA TTAGGCTCTG GAAAGACGCT CGTG 44

(2) INFORMATION FOR SEQ ID NO:116:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

AACGAGCGTC TTTG 14

(2) INFORMATION FOR SEQ ID NO:117:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

AACGAGCGTC ATTG 14

(2) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

TTTTTTTTTA ATTAGGCTCT GGAAAGACGC TCGTGAAACG AGCGTCTTTG 50

(2) INFORMATION FOR SEQ ID NO:119:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

TTTTCCAGAG CCTAATG 17

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position has a TET label."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

CCGGTCGTCC TGG 13

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

CAATTCCGGT GTACTCACCG GTTCC 25

(2) INFORMATION FOR SEQ ID NO:122:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position has a TET label."
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

CCGGTCGTCC TGGCAA 16

(2) INFORMATION FOR SEQ ID NO:123:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

TGTTTTGACC TCCATAGAAG ACCCTATAGT GAGTCGTATT AATTTCG 47

(2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

CGAAATTAAT ACGACTCACT ATA 23

(2) INFORMATION FOR SEQ ID NO:125:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

CGAAATTAAT ACGACTCACT ATACCCAGAA 30

(2) INFORMATION FOR SEQ ID NO:126:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

CGAAATTAAT ACGACT 16

(2) INFORMATION FOR SEQ ID NO:127:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

CGAAATTAAT ACG 13

(2) INFORMATION FOR SEQ ID NO:128:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

CGAAATTAAT AC 12

(2) INFORMATION FOR SEQ ID NO:129:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

CACTATAGGG TCTTCTATGG AGGTC 25

(2) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

ACTCACTATA GGGTCTTCTA TGGAGGTC 28

(2) INFORMATION FOR SEQ ID NO:131:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

GACTCACTAT AGGGTCTTCT ATGGAGGTC 29

(2) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

CGAAATTAAT ACGCAGTATG TTAGCAAACG 30

(2) INFORMATION FOR SEQ ID NO:133:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

GAACTGGCAT GATTAAGACT CCTTATTACC 30

(2) INFORMATION FOR SEQ ID NO:134:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

GAACTGGCAT GATTAAGACT CCTTATTAA 29

(2) INFORMATION FOR SEQ ID NO:135:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 326 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

```
Met Gly Val Gln Phe Gly Asp Phe Ile Pro Lys Asn Ile Ile Ser Phe
1               5                   10                  15

Glu Asp Leu Lys Gly Lys Lys Val Ala Ile Asp Gly Met Asn Ala Leu
            20                  25                  30

Tyr Gln Phe Leu Thr Ser Ile Arg Leu Arg Asp Gly Ser Pro Leu Arg
        35                  40                  45

Asn Arg Lys Gly Glu Ile Thr Ser Ala Tyr Asn Gly Val Phe Tyr Lys
    50                  55                  60

Thr Ile His Leu Leu Glu Asn Asp Ile Thr Pro Ile Trp Val Phe Asp
65                  70                  75                  80

Gly Glu Pro Pro Lys Leu Lys Glu Lys Thr Arg Lys Val Arg Arg Glu
                85                  90                  95

Met Lys Glu Lys Ala Glu Leu Lys Met Lys Glu Ala Ile Lys Lys Glu
            100                 105                 110

Asp Phe Glu Glu Ala Ala Lys Tyr Ala Lys Arg Val Ser Tyr Leu Thr
        115                 120                 125

Pro Lys Met Val Glu Asn Cys Lys Tyr Leu Leu Ser Leu Met Gly Ile
    130                 135                 140

Pro Tyr Val Glu Ala Pro Ser Glu Gly Glu Ala Gln Ala Ser Tyr Met
145                 150                 155                 160

Ala Lys Lys Gly Asp Val Trp Ala Val Val Ser Gln Asp Tyr Asp Ala
                165                 170                 175

Leu Leu Tyr Gly Ala Pro Arg Val Val Arg Asn Leu Thr Thr Thr Lys
            180                 185                 190

Glu Met Pro Glu Leu Ile Glu Leu Asn Glu Val Leu Glu Asp Leu Arg
        195                 200                 205

Ile Ser Leu Asp Asp Leu Ile Asp Ile Ala Ile Phe Met Gly Thr Asp
    210                 215                 220

Tyr Asn Pro Gly Gly Val Lys Gly Ile Gly Phe Lys Arg Ala Tyr Glu
225                 230                 235                 240

Leu Val Arg Ser Gly Val Ala Lys Asp Val Leu Lys Lys Glu Val Glu
                245                 250                 255
```

Tyr Tyr Asp Glu Ile Lys Arg Ile Phe Lys Glu Pro Lys Val Thr Asp
260 265 270

Asn Tyr Ser Leu Ser Leu Lys Leu Pro Asp Lys Glu Gly Ile Ile Lys
275 280 285

Phe Leu Val Asp Glu Asn Asp Phe Asn Tyr Asp Arg Val Lys Lys His
290 295 300

Val Asp Lys Leu Tyr Asn Leu Ile Ala Asn Lys Thr Lys Gln Lys Thr
305 310 315 320

Leu Asp Ala Trp Phe Lys
325

(2) INFORMATION FOR SEQ ID NO:136:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 340 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

Met Gly Val Pro Ile Gly Glu Ile Ile Pro Arg Lys Glu Ile Glu Leu
1 5 10 15

Glu Asn Leu Tyr Gly Lys Lys Ile Ala Ile Asp Ala Leu Asn Ala Ile
20 25 30

Tyr Gln Phe Leu Ser Thr Ile Arg Gln Lys Asp Gly Thr Pro Leu Met
35 40 45

Asp Ser Lys Gly Arg Ile Thr Ser His Leu Ser Gly Leu Phe Tyr Arg
50 55 60

Thr Ile Asn Leu Met Glu Ala Gly Ile Lys Pro Val Tyr Val Phe Asp
65 70 75 80

Gly Glu Pro Pro Glu Phe Lys Lys Lys Glu Leu Glu Lys Arg Arg Glu
85 90 95

Ala Arg Glu Glu Ala Glu Glu Lys Trp Arg Glu Ala Leu Glu Lys Gly
100 105 110

Glu Ile Glu Glu Ala Arg Lys Tyr Ala Gln Arg Ala Thr Arg Val Asn
115 120 125

Glu Met Leu Ile Glu Asp Ala Lys Lys Leu Leu Glu Leu Met Gly Ile
130 135 140

Pro Ile Val Gln Ala Pro Ser Glu Gly Glu Ala Gln Ala Ala Tyr Met
145 150 155 160

Ala Ala Lys Gly Ser Val Tyr Ala Ser Ala Ser Gln Asp Tyr Asp Ser
165 170 175

Leu Leu Phe Gly Ala Pro Arg Leu Val Arg Asn Leu Thr Ile Thr Gly
180 185 190

Lys Arg Lys Leu Pro Gly Lys Asn Val Tyr Val Glu Ile Lys Pro Glu
195 200 205

```
Leu Ile Ile Leu Glu Glu Val Leu Lys Glu Leu Lys Leu Thr Arg Glu
    210                 215                 220

Lys Leu Ile Glu Leu Ala Ile Leu Val Gly Thr Asp Tyr Asn Pro Gly
225                 230                 235                 240

Gly Ile Lys Gly Ile Gly Leu Lys Lys Ala Leu Glu Ile Val Arg His
                245                 250                 255

Ser Lys Asp Pro Leu Ala Lys Phe Gln Lys Gln Ser Asp Val Asp Leu
            260                 265                 270

Tyr Ala Ile Lys Glu Phe Phe Leu Asn Pro Pro Val Thr Asp Asn Tyr
        275                 280                 285

Asn Leu Val Trp Arg Asp Pro Asp Glu Glu Gly Ile Leu Lys Phe Leu
    290                 295                 300

Cys Asp Glu His Asp Phe Ser Glu Glu Arg Val Lys Asn Gly Leu Glu
305                 310                 315                 320

Arg Leu Lys Lys Ala Ile Lys Ser Gly Lys Gln Ser Thr Leu Glu Ser
                325                 330                 335

Trp Phe Lys Arg
            340
```

(2) INFORMATION FOR SEQ ID NO:137:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 380 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

```
Met Gly Ile Gln Gly Leu Ala Lys Leu Ile Ala Asp Val Ala Pro Ser
1               5                   10                  15

Ala Ile Arg Glu Asn Asp Ile Lys Ser Tyr Phe Gly Arg Lys Val Ala
            20                  25                  30

Ile Asp Ala Ser Met Ser Ile Tyr Gln Phe Leu Ile Ala Val Arg Gln
        35                  40                  45

Gly Gly Asp Val Leu Gln Asn Glu Glu Gly Glu Thr Thr Ser His Leu
    50                  55                  60

Met Gly Met Phe Tyr Arg Thr Ile Arg Met Met Glu Asn Gly Ile Lys
65                  70                  75                  80

Pro Val Tyr Val Phe Asp Gly Lys Pro Pro Gln Leu Lys Ser Gly Glu
                85                  90                  95

Leu Ala Lys Arg Ser Glu Arg Arg Ala Glu Ala Glu Lys Gln Leu Gln
            100                 105                 110

Gln Ala Gln Ala Ala Gly Ala Glu Gln Glu Val Glu Lys Phe Thr Lys
        115                 120                 125

Arg Leu Val Lys Val Thr Lys Gln His Asn Asp Glu Cys Lys His Leu
    130                 135                 140
```

Leu Ser Leu Met Gly Ile Pro Tyr Leu Asp Ala Pro Ser Glu Ala Glu
145 150 155 160

Ala Ser Cys Ala Ala Leu Val Lys Ala Gly Lys Val Tyr Ala Ala Ala
165 170 175

Thr Glu Asp Met Asp Cys Leu Thr Phe Gly Ser Pro Val Leu Met Arg
180 185 190

His Leu Thr Ala Ser Glu Ala Lys Lys Leu Pro Ile Gln Glu Phe His
195 200 205

Leu Ser Arg Ile Leu Gln Glu Leu Gly Leu Asn Gln Glu Gln Phe Val
210 215 220

Asp Leu Cys Ile Leu Leu Gly Ser Asp Tyr Cys Glu Ser Ile Arg Gly
225 230 235 240

Ile Gly Pro Lys Arg Ala Val Asp Leu Ile Gln Lys His Lys Ser Ile
245 250 255

Glu Glu Ile Val Arg Arg Leu Asp Pro Asn Lys Tyr Pro Val Pro Glu
260 265 270

Asn Trp Leu His Lys Glu Ala His Gln Leu Phe Leu Glu Pro Glu Val
275 280 285

Leu Asp Pro Glu Ser Val Glu Leu Lys Trp Ser Glu Pro Asn Glu Glu
290 295 300

Glu Leu Ile Lys Phe Met Cys Gly Glu Lys Gln Phe Ser Glu Glu Arg
305 310 315 320

Ile Arg Ser Gly Val Lys Arg Leu Ser Lys Ser Arg Gln Gly Ser Thr
325 330 335

Gln Gly Arg Leu Asp Asp Phe Phe Lys Val Thr Gly Ser Leu Ser Ser
340 345 350

Ala Lys Arg Lys Glu Pro Glu Pro Lys Gly Ser Thr Lys Lys Lys Ala
355 360 365

Lys Thr Gly Ala Ala Gly Lys Phe Lys Arg Gly Lys
370 375 380

(2) INFORMATION FOR SEQ ID NO:138:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 378 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

Met Gly Ile His Gly Leu Ala Lys Leu Ile Ala Asp Val Ala Pro Ser
1 5 10 15

Ala Ile Arg Glu Asn Asp Ile Lys Ser Tyr Phe Gly Arg Lys Val Ala
20 25 30

Ile Asp Ala Ser Met Ser Ile Tyr Gln Phe Leu Ile Ala Val Arg Gln
35 40 45

```
Gly Gly Asp Val Leu Gln Asn Glu Glu Gly Glu Thr Thr Ser Leu Met
     50                  55                  60

Gly Met Phe Tyr Arg Thr Ile Arg Met Glu Asn Gly Ile Lys Pro Val
65                  70                  75                  80

Tyr Val Phe Asp Gly Lys Pro Pro Gln Leu Lys Ser Gly Glu Leu Ala
                85                  90                  95

Lys Arg Ser Glu Arg Arg Ala Glu Ala Glu Lys Gln Leu Gln Gln Ala
            100                 105                 110

Gln Glu Ala Gly Met Glu Glu Glu Val Glu Lys Phe Thr Lys Arg Leu
        115                 120                 125

Val Lys Val Thr Lys Gln His Asn Asp Glu Cys Lys His Leu Leu Ser
    130                 135                 140

Leu Met Gly Ile Pro Tyr Leu Asp Ala Pro Ser Glu Ala Glu Ala Ser
145                 150                 155                 160

Cys Ala Ala Leu Ala Lys Ala Gly Lys Val Tyr Ala Ala Ala Thr Glu
                165                 170                 175

Asp Met Asp Cys Leu Thr Phe Gly Ser Pro Val Leu Met Arg His Leu
            180                 185                 190

Thr Ala Ser Glu Ala Lys Lys Leu Pro Ile Gln Glu Phe His Leu Ser
        195                 200                 205

Arg Val Leu Gln Glu Leu Gly Leu Asn Gln Glu Gln Phe Val Asp Leu
    210                 215                 220

Cys Ile Leu Leu Gly Ser Asp Tyr Cys Glu Ser Ile Arg Gly Ile Gly
225                 230                 235                 240

Ala Lys Arg Ala Val Asp Leu Ile Gln Lys His Lys Ser Ile Glu Glu
                245                 250                 255

Ile Val Arg Arg Leu Asp Pro Ser Lys Tyr Pro Val Pro Glu Asn Trp
            260                 265                 270

Leu His Lys Glu Ala Gln Gln Leu Phe Leu Glu Pro Glu Val Val Asp
        275                 280                 285

Pro Glu Ser Val Glu Leu Lys Trp Ser Glu Pro Asn Glu Glu Glu Leu
    290                 295                 300

Val Lys Phe Met Cys Gly Glu Lys Gln Phe Ser Glu Glu Arg Ile Arg
305                 310                 315                 320

Ser Gly Val Lys Arg Leu Ser Lys Ser Arg Gln Gly Ser Thr Gln Gly
                325                 330                 335

Arg Leu Asp Asp Phe Phe Lys Val Thr Gly Ser Leu Ser Ser Ala Lys
            340                 345                 350

Arg Lys Glu Pro Glu Pro Lys Gly Pro Ala Lys Lys Lys Ala Lys Thr
        355                 360                 365

Gly Gly Ala Gly Lys Phe Arg Arg Gly Lys
    370                 375
```

(2) INFORMATION FOR SEQ ID NO:139:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 382 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

```
Met Gly Ile Lys Gly Leu Asn Ala Ile Ile Ser Glu His Val Pro Ser
1               5                   10                  15

Ala Ile Arg Lys Ser Asp Ile Lys Ser Phe Phe Gly Arg Lys Val Ala
            20                  25                  30

Ile Asp Ala Ser Met Ser Leu Tyr Gln Phe Leu Ile Ala Val Arg Gln
        35                  40                  45

Gln Asp Gly Gly Gln Leu Thr Asn Glu Ala Gly Glu Thr Thr Ser His
    50                  55                  60

Leu Met Gly Met Phe Tyr Arg Thr Leu Arg Met Ile Asp Asn Gly Ile
65                  70                  75                  80

Lys Pro Cys Tyr Val Phe Asp Gly Lys Pro Pro Asp Leu Lys Ser His
                85                  90                  95

Glu Leu Thr Lys Arg Ser Ser Arg Arg Val Glu Thr Glu Lys Lys Leu
            100                 105                 110

Ala Glu Ala Thr Thr Glu Leu Glu Lys Met Lys Gln Glu Arg Arg Leu
        115                 120                 125

Val Lys Val Ser Lys Glu His Asn Glu Glu Ala Gln Lys Leu Leu Gly
    130                 135                 140

Leu Met Gly Ile Pro Tyr Ile Ile Ala Pro Thr Glu Ala Glu Ala Gln
145                 150                 155                 160

Cys Ala Glu Leu Ala Lys Lys Gly Lys Val Tyr Ala Ala Ala Ser Glu
                165                 170                 175

Asp Met Asp Thr Leu Cys Tyr Arg Thr Pro Phe Leu Leu Arg His Leu
            180                 185                 190

Thr Phe Ser Glu Ala Lys Lys Glu Pro Ile His Glu Ile Asp Thr Glu
        195                 200                 205

Leu Val Leu Arg Gly Leu Asp Leu Thr Ile Glu Gln Phe Val Asp Leu
    210                 215                 220

Cys Ile Met Leu Gly Cys Asp Tyr Cys Glu Ser Ile Arg Gly Val Gly
225                 230                 235                 240

Pro Val Thr Ala Leu Lys Leu Ile Lys Thr His Gly Ser Ile Glu Lys
                245                 250                 255

Ile Val Glu Phe Ile Glu Ser Gly Glu Ser Asn Asn Thr Lys Trp Lys
            260                 265                 270

Ile Pro Glu Asp Trp Pro Tyr Lys Gln Ala Arg Met Leu Phe Leu Asp
        275                 280                 285

Pro Glu Val Ile Asp Gly Asn Glu Ile Asn Leu Lys Trp Ser Pro Pro
    290                 295                 300
```

```
Lys Glu Lys Glu Leu Ile Glu Tyr Leu Cys Asp Asp Lys Lys Phe Ser
305                 310                 315                 320

Glu Glu Arg Val Lys Ser Gly Ile Ser Arg Leu Lys Lys Gly Leu Lys
                325                 330                 335

Ser Gly Ile Gln Gly Arg Leu Asp Gly Phe Phe Gln Val Val Pro Lys
            340                 345                 350

Thr Lys Glu Gln Leu Ala Ala Ala Ala Lys Arg Ala Gln Glu Asn Lys
        355                 360                 365

Lys Leu Asn Lys Asn Lys Asn Lys Val Thr Lys Gly Arg Arg
    370                 375                 380
```

(2) INFORMATION FOR SEQ ID NO:140:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 387 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

```
Met Gly Val His Ser Phe Trp Asp Ile Ala Gly Pro Thr Ala Arg Pro
1               5                   10                  15

Val Arg Leu Glu Ser Leu Glu Asp Lys Arg Met Ala Val Asp Ala Ser
            20                  25                  30

Ile Trp Ile Tyr Gln Phe Leu Lys Ala Val Arg Asp Gln Glu Gly Asn
        35                  40                  45

Ala Val Lys Asn Ser His Ile Thr Gly Phe Phe Arg Arg Ile Cys Lys
    50                  55                  60

Leu Leu Tyr Phe Gly Ile Arg Pro Val Phe Val Phe Asp Gly Gly Val
65                  70                  75                  80

Pro Val Leu Lys Arg Glu Thr Ile Arg Gln Arg Lys Glu Arg Arg Gln
                85                  90                  95

Gly Lys Arg Glu Ser Ala Lys Ser Thr Ala Arg Lys Leu Leu Ala Leu
            100                 105                 110

Gln Leu Gln Asn Gly Ser Asn Asp Asn Glu Val Thr Met Asp Met Ile
        115                 120                 125

Lys Glu Val Gln Glu Leu Leu Ser Arg Phe Gly Ile Pro Tyr Ile Thr
    130                 135                 140

Ala Pro Met Glu Ala Glu Ala Gln Cys Ala Glu Leu Leu Gln Leu Asn
145                 150                 155                 160

Leu Val Asp Gly Ile Ile Thr Asp Asp Ser Asp Val Phe Leu Phe Gly
                165                 170                 175

Gly Thr Lys Ile Tyr Lys Asn Met Phe His Glu Lys Asn Tyr Val Glu
            180                 185                 190

Phe Tyr Asp Ala Glu Ser Ile Leu Lys Leu Leu Gly Leu Asp Arg Lys
        195                 200                 205
```

Asn Met Ile Glu Leu Ala Gln Leu Leu Gly Ser Asp Tyr Thr Asn Gly
    210   215   220

Leu Lys Gly Met Gly Pro Val Ser Ser Ile Glu Val Ile Ala Glu Phe
225   230   235   240

Gly Asn Leu Lys Asn Phe Lys Asp Trp Tyr Asn Asn Gly Gln Phe Asp
                245   250   255

Lys Arg Lys Gln Glu Thr Glu Asn Lys Phe Glu Lys Asp Leu Arg Lys
            260   265   270

Lys Leu Val Asn Asn Glu Ile Ile Leu Asp Asp Asp Phe Pro Ser Val
        275   280   285

Met Val Tyr Asp Ala Tyr Met Arg Pro Glu Val Asp His Asp Thr Thr
    290   295   300

Pro Phe Val Trp Gly Val Pro Asp Leu Asp Met Leu Arg Ser Phe Met
305   310   315   320

Lys Thr Gln Leu Gly Trp Pro His Glu Lys Ser Asp Glu Ile Leu Ile
                325   330   335

Pro Leu Ile Arg Asp Val Asn Lys Arg Lys Lys Lys Gly Lys Gln Lys
            340   345   350

Arg Ile Asn Glu Phe Phe Pro Arg Glu Tyr Ile Ser Gly Asp Lys Lys
        355   360   365

Leu Asn Thr Ser Lys Arg Ile Ser Thr Ala Thr Gly Lys Leu Lys Lys
    370   375   380

Arg Lys Met
385

(2) INFORMATION FOR SEQ ID NO:141:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 488 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

Met Gly Val Ser Gly Leu Trp Asn Ile Leu Glu Pro Val Lys Arg Pro
1   5   10   15

Val Lys Leu Glu Thr Leu Val Asn Lys Arg Leu Ala Ile Asp Ala Ser
            20   25   30

Ile Trp Ile Tyr Gln Phe Leu Lys Ala Val Arg Asp Lys Glu Gly Asn
        35   40   45

Gln Leu Lys Ser Ser His Val Val Gly Phe Phe Arg Arg Ile Cys Lys
    50   55   60

Leu Leu Phe Phe Gly Ile Lys Pro Val Phe Val Phe Asp Gly Gly Ala
65   70   75   80

Pro Ser Leu Lys Arg Gln Thr Ile Gln Lys Arg Gln Ala Arg Arg Leu
                85   90   95

Asp Arg Glu Glu Asn Ala Thr Val Thr Ala Asn Lys Leu Leu Ala Leu
100 105 110

Gln Met Arg His Gln Ala Met Leu Leu Lys Arg Asp Ala Asp Glu Val
115 120 125

Thr Gln Val Met Ile Lys Glu Cys Gln Glu Leu Leu Arg Leu Phe Gly
130 135 140

Leu Pro Tyr Ile Val Ala Pro Gln Glu Ala Glu Ala Gln Cys Ser Lys
145 150 155 160

Leu Leu Glu Leu Lys Leu Val Asp Gly Ile Val Thr Asp Asp Ser Asp
165 170 175

Val Phe Leu Phe Gly Gly Thr Arg Val Tyr Arg Asn Met Phe Asn Gln
180 185 190

Asn Lys Phe Val Glu Leu Tyr Leu Met Asp Asp Met Lys Arg Glu Phe
195 200 205

Asn Val Asn Gln Met Asp Leu Ile Lys Leu Ala His Leu Leu Gly Ser
210 215 220

Asp Tyr Thr Met Gly Leu Ser Arg Val Gly Pro Val Leu Ala Leu Glu
225 230 235 240

Ile Leu His Glu Phe Pro Gly Asp Thr Gly Leu Phe Glu Phe Lys Lys
245 250 255

Trp Phe Gln Arg Leu Ser Thr Gly His Ala Ser Lys Asn Asp Val Asn
260 265 270

Thr Pro Val Lys Lys Arg Ile Asn Lys Leu Val Gly Lys Ile Ile Leu
275 280 285

Pro Ser Glu Phe Pro Asn Pro Leu Val Asp Glu Ala Tyr Leu His Pro
290 295 300

Ala Val Asp Asp Ser Lys Gln Ser Phe Gln Trp Gly Ile Pro Asp Leu
305 310 315 320

Asp Glu Leu Arg Gln Phe Leu Met Ala Thr Val Gly Trp Ser Lys Gln
325 330 335

Arg Thr Asn Glu Val Leu Leu Pro Val Ile Gln Asp Met His Lys Lys
340 345 350

Gln Phe Val Gly Thr Gln Ser Asn Leu Thr Gln Phe Phe Glu Gly Gly
355 360 365

Asn Thr Asn Val Tyr Ala Pro Arg Val Ala Tyr His Phe Lys Ser Lys
370 375 380

Arg Leu Glu Asn Ala Leu Ser Ser Phe Lys Asn Gln Ile Ser Asn Gln
385 390 395 400

Ser Pro Met Ser Glu Glu Ile Gln Ala Asp Ala Asp Ala Phe Gly Glu
405 410 415

Ser Lys Gly Ser Asp Glu Leu Gln Ser Arg Ile Leu Arg Arg Lys Lys
420 425 430

Met Met Ala Ser Lys Asn Ser Ser Asp Ser Asp Ser Asp Ser Glu Asp
435 440 445

Asn Phe Leu Ala Ser Leu Thr Pro Lys Thr Asn Ser Ser Ser Ile Ser
450 455 460

Ile Glu Asn Leu Pro Arg Lys Thr Lys Leu Ser Thr Ser Leu Leu Lys
465 470 475 480

Lys Pro Ser Lys Arg Arg Arg Lys
485

(2) INFORMATION FOR SEQ ID NO:142:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 550 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

Met Gly Val Gln Gly Leu Trp Lys Leu Leu Glu Cys Ser Gly Arg Gln
1 5 10 15

Val Ser Pro Glu Ala Leu Glu Gly Lys Ile Leu Ala Val Asp Ile Ser
20 25 30

Ile Trp Leu Asn Gln Ala Leu Lys Gly Val Arg Asp Arg His Gly Asn
35 40 45

Ser Ile Glu Asn Pro His Leu Leu Thr Leu Phe His Arg Leu Cys Lys
50 55 60

Leu Leu Phe Phe Arg Ile Arg Pro Ile Phe Val Phe Asp Gly Asp Ala
65 70 75 80

Pro Leu Leu Lys Lys Gln Thr Leu Val Lys Arg Arg Gln Arg Lys Asp
85 90 95

Leu Ala Ser Ser Asp Ser Arg Lys Thr Thr Glu Lys Leu Leu Lys Thr
100 105 110

Phe Leu Lys Arg Gln Ala Ile Lys Thr Glu Arg Ile Ala Ala Thr Val
115 120 125

Thr Gly Gln Met Phe Leu Glu Ser Gln Glu Leu Leu Arg Leu Phe Gly
130 135 140

Ile Pro Tyr Ile Gln Ala Pro Met Glu Ala Glu Ala Gln Cys Ala Ile
145 150 155 160

Leu Asp Leu Thr Asp Gln Thr Ser Gly Thr Ile Thr Asp Asp Ser Asp
165 170 175

Ile Trp Leu Phe Gly Ala Arg His Val Tyr Arg Asn Phe Phe Asn Lys
180 185 190

Asn Lys Phe Val Glu Tyr Tyr Gln Tyr Val Asp Phe His Asn Gln Leu
195 200 205

Gly Leu Asp Arg Asn Lys Leu Ile Asn Leu Ala Tyr Leu Leu Gly Ser
210 215 220

Asp Tyr Thr Glu Gly Ile Pro Thr Val Gly Cys Val Thr Ala Met Glu
225 230 235 240

```
Ile Leu Asn Glu Phe Pro Gly His Gly Leu Glu Pro Leu Leu Lys Phe
                245                 250                 255

Ser Glu Trp Trp His Glu Ala Gln Lys Asn Pro Lys Ile Arg Pro Asn
            260                 265                 270

Pro His Asp Thr Lys Val Lys Lys Lys Leu Arg Thr Leu Gln Leu Thr
        275                 280                 285

Pro Gly Phe Pro Asn Pro Ala Val Ala Glu Ala Tyr Leu Lys Pro Val
    290                 295                 300

Val Asp Asp Ser Lys Gly Ser Phe Leu Trp Gly Lys Pro Asp Leu Asp
305                 310                 315                 320

Lys Ile Arg Glu Phe Cys Gln Arg Tyr Phe Gly Trp Asn Arg Thr Lys
                325                 330                 335

Thr Asp Glu Ser Leu Phe Pro Val Leu Lys Gln Leu Asp Ala Gln Gln
            340                 345                 350

Thr Gln Leu Arg Ile Asp Ser Phe Phe Arg Leu Ala Gln Gln Glu Lys
        355                 360                 365

Glu Asp Ala Lys Arg Ile Lys Ser Gln Arg Leu Asn Arg Ala Val Thr
    370                 375                 380

Cys Met Leu Arg Lys Glu Lys Glu Ala Ala Ala Ser Glu Ile Glu Ala
385                 390                 395                 400

Val Ser Val Ala Met Glu Lys Glu Phe Glu Leu Leu Asp Lys Ala Lys
                405                 410                 415

Arg Lys Thr Gln Lys Arg Gly Ile Thr Asn Thr Leu Glu Glu Ser Ser
            420                 425                 430

Ser Leu Lys Arg Lys Arg Leu Ser Asp Ser Lys Arg Lys Asn Thr Cys
        435                 440                 445

Gly Gly Phe Leu Gly Glu Thr Cys Leu Ser Glu Ser Ser Asp Gly Ser
    450                 455                 460

Ser Ser Glu His Ala Glu Ser Ser Ser Leu Met Asn Val Gln Arg Arg
465                 470                 475                 480

Thr Ala Ala Lys Glu Pro Lys Thr Ser Ala Ser Asp Ser Gln Asn Ser
                485                 490                 495

Val Lys Glu Ala Pro Val Lys Asn Gly Gly Ala Thr Thr Ser Ser Ser
            500                 505                 510

Ser Asp Ser Asp Asp Asp Gly Gly Lys Glu Lys Met Val Leu Val Thr
        515                 520                 525

Ala Arg Ser Val Phe Gly Lys Lys Arg Arg Lys Leu Arg Arg Ala Arg
    530                 535                 540

Gly Arg Lys Arg Lys Thr
545                 550
```

(2) INFORMATION FOR SEQ ID NO:143:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 543 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

Met Gly Val Gln Gly Leu Trp Lys Leu Leu Glu Cys Ser Gly His Arg
1   5   10   15

Val Ser Pro Glu Ala Leu Glu Gly Lys Val Leu Ala Val Asp Ile Ser
            20   25   30

Ile Trp Leu Asn Gln Ala Leu Lys Gly Val Arg Asp Ser His Gly Asn
        35   40   45

Val Ile Glu Asn Ala His Leu Leu Thr Leu Phe His Arg Leu Cys Lys
    50   55   60

Leu Leu Phe Phe Arg Ile Arg Pro Ile Phe Val Phe Asp Gly Asp Ala
65   70   75   80

Pro Leu Leu Lys Lys Gln Thr Leu Ala Lys Arg Arg Gln Arg Lys Asp
                85   90   95

Ser Ala Ser Ile Asp Ser Arg Lys Thr Thr Glu Lys Leu Leu Lys Thr
            100   105   110

Phe Leu Lys Arg Gln Ala Leu Lys Thr Asp Arg Ile Ala Ala Ser Val
        115   120   125

Thr Gly Gln Met Phe Leu Glu Ser Gln Glu Leu Leu Arg Leu Phe Gly
    130   135   140

Val Pro Tyr Ile Gln Ala Pro Met Glu Ala Glu Ala Gln Cys Ala Val
145   150   155   160

Leu Asp Leu Ser Asp Gln Thr Ser Gly Thr Ile Thr Asp Asp Ser Asp
                165   170   175

Ile Trp Leu Phe Gly Ala Arg His Val Tyr Lys Asn Phe Phe Asn Lys
            180   185   190

Asn Lys Phe Val Glu Tyr Tyr Gln Tyr Val Asp Phe Tyr Ser Gln Leu
        195   200   205

Gly Leu Asp Arg Asn Lys Leu Ile Asn Leu Ala Tyr Leu Leu Gly Ser
    210   215   220

Asp Tyr Thr Glu Gly Ile Pro Thr Val Gly Cys Val Thr Ala Met Glu
225   230   235   240

Ile Leu Asn Glu Phe Pro Gly Arg Gly Leu Asp Pro Leu Leu Lys Phe
                245   250   255

Ser Glu Trp Trp His Glu Ala Gln Asn Asn Lys Lys Val Ala Glu Asn
            260   265   270

Pro Tyr Asp Thr Lys Val Lys Lys Lys Leu Arg Lys Leu Gln Leu Thr
        275   280   285

Pro Gly Phe Pro Asn Pro Ala Val Ala Asp Ala Tyr Leu Arg Pro Val
    290   295   300

Val Asp Asp Ser Arg Gly Ser Phe Leu Trp Gly Lys Pro Asp Val Asp
305   310   315   320

Lys Ile Arg Glu Phe Cys Gln Arg Tyr Phe Gly Trp Asn Arg Met Lys
                325   330   335

Thr Asp Glu Ser Leu Tyr Pro Val Leu Lys His Leu Asn Ala His Gln
340 345 350

Thr Gln Leu Arg Ile Asp Ser Phe Phe Arg Leu Ala Gln Gln Glu Lys
355 360 365

Gln Asp Ala Lys Leu Ile Lys Ser His Arg Leu Ser Arg Ala Val Thr
370 375 380

Cys Met Leu Arg Lys Glu Arg Glu Glu Lys Ala Pro Glu Leu Thr Lys
385 390 395 400

Val Thr Glu Ala Met Glu Lys Glu Phe Glu Leu Leu Asp Asp Ala Lys
405 410 415

Gly Lys Thr Gln Lys Arg Glu Leu Pro Tyr Lys Lys Glu Thr Ser Val
420 425 430

Pro Lys Arg Arg Arg Pro Ser Gly Asn Gly Gly Phe Leu Gly Asp Pro
435 440 445

Tyr Cys Ser Glu Ser Pro Gln Glu Ser Ser Cys Glu Asp Gly Glu Gly
450 455 460

Ser Ser Val Met Ser Ala Arg Gln Arg Ser Ala Ala Glu Ser Ser Lys
465 470 475 480

Ile Gly Cys Ser Asp Val Pro Asp Leu Val Arg Asp Ser Pro His Gly
485 490 495

Arg Gln Gly Cys Val Ser Thr Ser Ser Ser Asp Ser Glu Asp Gly Glu
500 505 510

Asp Lys Ala Lys Thr Val Leu Val Thr Ala Arg Pro Val Phe Gly Lys
515 520 525

Lys Arg Arg Lys Leu Lys Ser Met Lys Arg Arg Lys Lys Lys Thr
530 535 540

(2) INFORMATION FOR SEQ ID NO:144:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 527 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

Met Gly Val Gln Gly Leu Trp Lys Leu Leu Glu Cys Ser Gly Arg Pro
1 5 10 15

Ile Asn Pro Gly Thr Leu Glu Gly Lys Ile Leu Ala Val Asp Ile Ser
20 25 30

Ile Trp Leu Asn Gln Ala Val Lys Gly Ala Arg Asp Arg Gln Gly Asn
35 40 45

Ala Ile Gln Asn Ala His Leu Leu Thr Leu Phe His Arg Leu Cys Lys
50 55 60

Leu Leu Phe Phe Arg Ile Arg Pro Ile Phe Val Phe Asp Gly Glu Ala
65 70 75 80

Pro Leu Leu Lys Arg Gln Thr Leu Ala Lys Arg Arg Gln Arg Thr Asp
85 90 95

Lys Ala Ser Asn Asp Ala Arg Lys Thr Asn Glu Lys Leu Leu Arg Thr
100 105 110

Phe Leu Lys Arg Gln Ala Ile Lys Ala Glu Arg Ile Ala Ala Thr Val
115 120 125

Thr Gly Gln Met Cys Leu Glu Ser Gln Glu Leu Leu Gln Leu Phe Gly
130 135 140

Ile Pro Tyr Ile Val Ala Pro Met Glu Ala Glu Ala Gln Cys Ala Ile
145 150 155 160

Leu Asp Leu Thr Asp Gln Thr Ser Gly Thr Ile Thr Asp Asp Ser Asp
165 170 175

Ile Trp Leu Phe Gly Ala Arg His Val Tyr Lys Asn Phe Phe Ser Gln
180 185 190

Asn Lys His Val Glu Tyr Tyr Gln Tyr Ala Asp Ile His Asn Gln Leu
195 200 205

Gly Leu Asp Arg Ser Lys Leu Ile Asn Leu Ala Tyr Leu Leu Gly Ser
210 215 220

Asp Tyr Thr Glu Gly Ile Pro Thr Val Gly Tyr Val Ser Ala Met Glu
225 230 235 240

Ile Leu Asn Glu Phe Pro Gly Gln Gly Leu Glu Pro Leu Val Lys Phe
245 250 255

Lys Glu Trp Trp Ser Glu Ala Gln Lys Asp Lys Lys Met Arg Pro Asn
260 265 270

Pro Asn Asp Thr Lys Val Lys Lys Lys Leu Arg Leu Leu Asp Leu Gln
275 280 285

Gln Ser Phe Pro Asn Pro Ala Val Ala Ser Ala Tyr Leu Lys Pro Val
290 295 300

Val Asp Glu Ser Lys Ser Ala Phe Ser Trp Gly Arg Pro Asp Leu Glu
305 310 315 320

Gln Ile Arg Glu Phe Cys Glu Ser Arg Phe Gly Trp Tyr Arg Leu Lys
325 330 335

Thr Asp Glu Val Leu Leu Pro Val Leu Lys Gln Leu Asn Ala Gln Gln
340 345 350

Thr Gln Leu Arg Ile Asp Ser Phe Phe Arg Leu Glu Gln His Glu Ala
355 360 365

Ala Gly Leu Lys Ser Gln Arg Leu Arg Arg Ala Val Thr Cys Met Lys
370 375 380

Arg Lys Glu Arg Asp Val Glu Ala Glu Glu Val Glu Ala Ala Val Ala
385 390 395 400

Val Met Glu Arg Glu Cys Thr Asn Gln Arg Lys Gly Gln Lys Thr Asn
405 410 415

Thr Lys Ser Gln Gly Thr Lys Arg Arg Lys Pro Thr Glu Cys Ser Gln
420 425 430

```
Glu Asp Gln Asp Pro Gly Gly Gly Phe Ile Gly Ile Glu Leu Lys Thr
        435                 440                 445

Leu Ser Ser Lys Ala Tyr Ser Ser Asp Gly Ser Ser Ser Asp Ala Glu
    450                 455                 460

Asp Leu Pro Ser Gly Leu Ile Asp Lys Gln Ser Gln Ser Gly Ile Val
465                 470                 475                 480

Gly Arg Gln Lys Ala Ser Asn Lys Val Glu Ser Ser Ser Ser Ser Asp
                485                 490                 495

Asp Glu Asp Arg Thr Val Met Val Thr Ala Lys Pro Val Phe Gln Gly
            500                 505                 510

Lys Lys Thr Lys Ser Lys Thr Met Lys Glu Thr Val Lys Arg Lys
        515                 520                 525
```

(2) INFORMATION FOR SEQ ID NO:145:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 434 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

```
Met Thr Ile Asn Gly Ile Trp Glu Trp Ala Asn His Val Val Arg Lys
1               5                   10                  15

Val Pro Asn Glu Thr Met Arg Asp Lys Thr Leu Ser Ile Asp Gly His
            20                  25                  30

Ile Trp Leu Tyr Glu Ser Leu Lys Gly Cys Glu Ala His His Gln Gln
        35                  40                  45

Thr Pro Asn Ser Tyr Leu Val Thr Phe Phe Thr Arg Ile Gln Arg Leu
    50                  55                  60

Leu Glu Leu Lys Ile Ile Pro Ile Val Val Phe Asp Asn Ile Asn Ala
65                  70                  75                  80

Ser Ser Ser Ala His Glu Ser Lys Asp Gln Asn Glu Phe Val Pro Arg
                85                  90                  95

Lys Arg Arg Ser Phe Gly Asp Ser Pro Phe Thr Asn Leu Val Asp His
            100                 105                 110

Val Tyr Lys Thr Asn Ala Leu Leu Thr Glu Leu Gly Ile Lys Val Ile
        115                 120                 125

Ile Ala Pro Gly Asp Gly Glu Ala Gln Cys Ala Arg Leu Glu Asp Leu
    130                 135                 140

Gly Val Thr Ser Gly Cys Ile Thr Thr Asp Phe Asp Tyr Phe Leu Phe
145                 150                 155                 160

Gly Gly Lys Asn Leu Tyr Arg Phe Asp Phe Thr Ala Gly Thr Ser Ser
                165                 170                 175

Thr Ala Cys Leu His Asp Ile Met His Leu Ser Leu Gly Arg Met Phe
            180                 185                 190
```

```
Met Glu Lys Lys Val Ser Arg Pro His Leu Ile Ser Thr Ala Ile Leu
        195                 200                 205

Leu Gly Cys Asp Tyr Phe Gln Arg Gly Val Gln Asn Ile Gly Ile Val
    210                 215                 220

Ser Val Phe Asp Ile Leu Gly Glu Phe Gly Asp Asp Gly Asn Glu Glu
225                 230                 235                     240

Ile Asp Pro His Val Ile Leu Asp Arg Phe Ala Ser Tyr Val Arg Glu
                245                 250                 255

Glu Ile Pro Ala Arg Ser Glu Asp Thr Gln Arg Lys Leu Arg Leu Arg
            260                 265                 270

Arg Lys Lys Tyr Asn Phe Pro Val Gly Phe Pro Asn Cys Asp Ala Val
        275                 280                 285

His Asn Ala Ile Thr Met Tyr Leu Arg Pro Pro Val Ser Ser Glu Ile
    290                 295                 300

Pro Lys Ile Ile Pro Arg Ala Ala Asn Phe Gln Gln Val Ala Glu Ile
305                 310                 315                     320

Met Met Lys Glu Cys Gly Trp Pro Ala Thr Arg Thr Gln Lys Glu Leu
                325                 330                 335

Ala Leu Ser Ile Arg Arg Lys Val His Leu Thr Thr Thr Val Ala Gln
            340                 345                 350

Thr Arg Ile Pro Asp Phe Phe Ala Ala Thr Lys Ser Lys Asn Phe Thr
        355                 360                 365

Pro Ile Val Glu Pro Cys Glu Ser Leu Glu Asp Tyr Ile Ser Ala Asn
    370                 375                 380

Asn Thr Trp Met Arg Lys Arg Lys Arg Ser Glu Ser Pro Gln Ile Leu
385                 390                 395                     400

Gln His His Ala Lys Arg Gln Val Pro Asp Arg Lys Arg Ser Val Lys
                405                 410                 415

Ile Arg Ala Phe Lys Pro Tyr Pro Thr Asp Val Ile Glu Leu Gly Asp
            420                 425                 430

Ser Asp
```

(2) INFORMATION FOR SEQ ID NO:146:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
- (A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

TACTGACTCA CTATAGGGTC TTCTATGGAG GTC 33

(2) INFORMATION FOR SEQ ID NO:147:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

TTTTTTTTTA ATTAGGCTCT GGAAGACGCT GAAAGCGTCT TG 42

(2) INFORMATION FOR SEQ ID NO:148:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

TTTTTTTTTA ATTAGGCTCT GGAAGACGGA ACGTCTTG 38

(2) INFORMATION FOR SEQ ID NO:149:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

TTTTTTTTTA ATTAGGCTCT GGAAGAGAAT CTTG 34

(2) INFORMATION FOR SEQ ID NO:150:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

TTTTTTTTTA ATTAGGCTCT GGAAGGAACT TG 32

(2) INFORMATION FOR SEQ ID NO:151:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

TTTTTTTTTA ATTAGGCTCT GGAAG 25

(2) INFORMATION FOR SEQ ID NO:152:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"
(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1
(D) OTHER INFORMATION: /note= "The residue at this position contains a TET-label."
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

ATTAGAAAGG AAGGGAAGAA AGCGAA 26

(2) INFORMATION FOR SEQ ID NO:153:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

ACGGGGAAAG CCGGCGAACG TGGCGAGAAA 30

(2) INFORMATION FOR SEQ ID NO:154:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

TGACGGGGAA AGCCGGCGAA CGTGGCGAGA 30

(2) INFORMATION FOR SEQ ID NO:155:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

CTTGACGGGG AAAGCCGGCG AACGTGGCGA 30

(2) INFORMATION FOR SEQ ID NO:156:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

GCTTGACGGG GAAAGCCGGC GAACGTGGCG 30

(2) INFORMATION FOR SEQ ID NO:157:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"
(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1
(D) OTHER INFORMATION: /note= "The residue as this position contains a fluoroscein label."
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:

AGAAAGGAAG GGAAGAAA 18

(2) INFORMATION FOR SEQ ID NO:158:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"
(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1
(D) OTHER INFORMATION: /note= "The residue at this position contains a fluoroscein label."
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:

TGGAGGTCAA AACATCGATA AGTCGAAGAA AGGAAGGGAA GAAAT 45

(2) INFORMATION FOR SEQ ID NO:159:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"
(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1
(D) OTHER INFORMATION: /note= "The residue at this position contains a fluoroscein label."
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:

TGTTTTGACC TCCA 14

(2) INFORMATION FOR SEQ ID NO:160:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:

ACACAGTGTC CTCCCGCTCC TCCTGAGCAA 30

(2) INFORMATION FOR SEQ ID NO:161:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"
(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1
(D) OTHER INFORMATION: /note= "The residue at this positions contains a fluoroscein label."
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:

TTTCCCTCCT CCTCTTCC 18

(2) INFORMATION FOR SEQ ID NO:162:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:

ATGAGGAAGA GGAGGAGGGT GCTCAGGAGG AGCGGGAGGA CACTGTGTCT GTCA 54

(2) INFORMATION FOR SEQ ID NO:163:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

TTCGCTTTCT TCCCTTCCTT TCTCGCCACG TTCGCCGGCT TTCCCCGTCA AGC 53

(2) INFORMATION FOR SEQ ID NO:164:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1011 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

ATGGGTGCGG ATATTGGTGA CCTCTTTGAG AGGGAAGAGG TCGAGCTTGA GTACTTCTCA 60

GGAAAGAAAA TTGCCGTTGA TGCTTTCAAC ACGCTATACC AGTTCATCTC GATAATAAGG 120

CAGCCTGACG GTACGCCGTT AAAGGACTCA CAGGGCAGAA TCACCTCTCA CCTTTCCGGA 180

ATCCTATACA GAGTCTCCAA CATGGTCGAG GTGGGAATCA GGCCGGTGTT TGTATTCGAC 240

GGAGAGCCAC CGGAGTTCAA GAAGGCTGAA ATTGAGGAGA GGAAAAAGAG AAGGGCTGAG 300

GCAGAGGAGA TGTGGATTGC GGCTTTGCAG GCAGGAGATA AGGACGCGAA AAAGTATGCT 360

CAGGCTGCAG GGAGGGTTGA CGAGTACATT GTTGACTCCG CAAAGACGCT TTTAAGTTAC 420

ATGGGGATTC CCTTTGTCGA TGCCCCGTCT GAAGGAGAGG CGCAGGCTGC TTACATGGCA 480

GCAAAAGGCG ATGTGGAGTA CACAGGAAGC CAGGATTACG ATTCTCTGCT CTTCGGAAGC 540

CCGAGACTCG CCAGAAATCT CGCAATAACG GGAAAAAGGA AGCTTCCCGG CAAAAATGTC 600

TATGTGGATG TAAAGCCGGA GATAATAATT CTGGAAAGCA ACCTCAAAAG GCTGGGTTTG 660

ACGAGGGAGC AGCTCATCGA CATAGCGATT CTGGTCGGGA CGGACTACAA TGAGGGTGTG 720

AAGGGTGTCG GCGTCAAGAA GGCTTTGAAC TACATCAAGA CCTACGGAGA TATTTTCAGG 780

GCACTCAAGG CTCTGAAAGT AAATATTGAC CACGTAGAGG AGATAAGGAA TTTCTTCCTG 840

AATCCTCCTG TGACTGACGA CTACAGAATA GAGTTCAGGG AGCCTGACTT TGAGAAGGCC 900

ATCGAGTTCC TGTGCGAGGA GCACGACTTC AGCAGGGAGA GGGTCGAGAA GGCCTTGGAG 960

AAGCTCAAAG CTCTGAAGTC AACCCAGGCC ACGCTTGAGA GGTGGTTCTG A 1011

(2) INFORMATION FOR SEQ ID NO:165:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 336 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

```
Met Gly Ala Asp Ile Gly Asp Leu Phe Glu Arg Glu Glu Val Glu Leu
1               5                   10                  15

Glu Tyr Phe Ser Gly Lys Lys Ile Ala Val Asp Ala Phe Asn Thr Leu
            20                  25                  30

Tyr Gln Phe Ile Ser Ile Ile Arg Gln Pro Asp Gly Thr Pro Leu Lys
        35                  40                  45

Asp Ser Gln Gly Arg Ile Thr Ser His Leu Ser Gly Ile Leu Tyr Arg
    50                  55                  60

Val Ser Asn Met Val Glu Val Gly Ile Arg Pro Val Phe Val Phe Asp
65                  70                  75                  80

Gly Glu Pro Pro Glu Phe Lys Lys Ala Glu Ile Glu Glu Arg Lys Lys
                85                  90                  95

Arg Arg Ala Glu Ala Glu Glu Met Trp Ile Ala Ala Leu Gln Ala Gly
            100                 105                 110

Asp Lys Asp Ala Lys Lys Tyr Ala Gln Ala Ala Gly Arg Val Asp Glu
        115                 120                 125

Tyr Ile Val Asp Ser Ala Lys Thr Leu Leu Ser Tyr Met Gly Ile Pro
    130                 135                 140

Phe Val Asp Ala Pro Ser Glu Gly Glu Ala Gln Ala Ala Tyr Met Ala
145                 150                 155                 160

Ala Lys Gly Asp Val Glu Tyr Thr Gly Ser Gln Asp Tyr Asp Ser Leu
                165                 170                 175

Leu Phe Gly Ser Pro Arg Leu Ala Arg Asn Leu Ala Ile Thr Gly Lys
            180                 185                 190

Arg Lys Leu Pro Gly Lys Asn Val Tyr Val Asp Val Lys Pro Glu Ile
        195                 200                 205

Ile Ile Leu Glu Ser Asn Leu Lys Arg Leu Gly Leu Thr Arg Glu Gln
    210                 215                 220

Leu Ile Asp Ile Ala Ile Leu Val Gly Thr Asp Tyr Asn Glu Gly Val
225                 230                 235                 240

Lys Gly Val Gly Val Lys Lys Ala Leu Asn Tyr Ile Lys Thr Tyr Gly
                245                 250                 255

Asp Ile Phe Arg Ala Leu Lys Ala Leu Lys Val Asn Ile Asp His Val
            260                 265                 270
```

```
Glu Glu Ile Arg Asn Phe Phe Leu Asn Pro Pro Val Thr Asp Asp Tyr
        275                 280                 285

Arg Ile Glu Phe Arg Glu Pro Asp Phe Glu Lys Ala Ile Glu Phe Leu
    290                 295                 300

Cys Glu Glu His Asp Phe Ser Arg Glu Arg Val Glu Lys Ala Leu Glu
305                 310                 315                 320

Lys Leu Lys Ala Leu Lys Ser Thr Gln Ala Thr Leu Glu Arg Trp Phe
                325                 330                 335
```

(2) INFORMATION FOR SEQ ID NO:166:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

CCGTCAACAT TTACCATGGG TGCGGA 26

(2) INFORMATION FOR SEQ ID NO:167:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

CCGCCACCTC GTAGTCGACA TCCTTTTCGT G 31

(2) INFORMATION FOR SEQ ID NO:168:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

GGCGACCACA CCCGTCCTGT 20

(2) INFORMATION FOR SEQ ID NO:169:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

CCACGATGCG TCCGGCGTAG 20

(2) INFORMATION FOR SEQ ID NO:170:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(29, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3' amine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

AACGAGGCGC ACCCACCCAA GGCACAGCN 29

(2) INFORMATION FOR SEQ ID NO:171:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

ACGGGTCAAT GTCCATGCCC CAAAGA 26

(2) INFORMATION FOR SEQ ID NO:172:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 23..27
(D) OTHER INFORMATION: /note= "The residues at these positions are 2' o-methyls"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(28, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3'amine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

GTCTGAGATG AAAGTGCGCC TCGTTAAN 28

(2) INFORMATION FOR SEQ ID NO:173:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a spacer containing a fluorescein label"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

NTCTTCGCAC ATTTCATCTC AGACGGA 27

(2) INFORMATION FOR SEQ ID NO:174:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1..21
(D) OTHER INFORMATION: /note= "The residues at these positions are 2' o-methyls"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(22, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3' amine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

GCTGTGCCTT GGGTGGGTGC GN 22

(2) INFORMATION FOR SEQ ID NO:175:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(29, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3' amine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

AACGAGGCGC ACCCACCCAA GGCACAGCN 29

(2) INFORMATION FOR SEQ ID NO:176:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

ACGGGTCAAT GTCCATGCCC CAAAGA 26

(2) INFORMATION FOR SEQ ID NO:177:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 23..27
(D) OTHER INFORMATION: /note= "The residues at these positions are 2' o-methyls"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(28, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3' amine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

GTCTGAGATG AAAGTGCGCC TCGTTAAN 28

(2) INFORMATION FOR SEQ ID NO:178:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"
(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a spacer containing a fluorescein label"
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

NTCTTCGCAC ATTTCATCTC AGAC 24

(2) INFORMATION FOR SEQ ID NO:179:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1..17
(D) OTHER INFORMATION: /note= "The residues at these positions are 2' o-methyls"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(18, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3' amine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

GCTGTGCCTT GGGTGGGN 18

(2) INFORMATION FOR SEQ ID NO:180:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1..19
(D) OTHER INFORMATION: /note= "The residues at these positions are 2' o-methyls"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(20, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3' amine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

GCTGTGCCTT GGGTGGGTGN 20

(2) INFORMATION FOR SEQ ID NO:181:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

```
    (ix) FEATURE:
           (A) NAME/KEY: modified_base
           (B) LOCATION: 1..21
           (D) OTHER INFORMATION: /note= "The residues at these
  positions are 2' o-methyls"

    (ix) FEATURE:
           (A) NAME/KEY: misc_difference
           (B) LOCATION: replace(22, "")
           (D) OTHER INFORMATION: /note= "The residue at this
  position is a 3' amine"

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:

GCTGTGCCTT GGGTGGGTGC GN                                             22

(2) INFORMATION FOR SEQ ID NO:182:

     (i) SEQUENCE CHARACTERISTICS:
           (A) LENGTH: 23 base pairs
           (B) TYPE: nucleic acid
           (C) STRANDEDNESS: single
           (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: other nucleic acid
           (A) DESCRIPTION: /desc = "DNA"

    (ix) FEATURE:
           (A) NAME/KEY: modified_base
           (B) LOCATION: 1..22
           (D) OTHER INFORMATION: /note= "The residues at these
  positions are 2' o-methyls"

    (ix) FEATURE:
           (A) NAME/KEY: misc_difference
           (B) LOCATION: replace(23, "")
           (D) OTHER INFORMATION: /note= "The residue at this
  position is a 3' amine"
    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

GCTGTGCCTT GGGTGGGTGC GCN                                            23

(2) INFORMATION FOR SEQ ID NO:183:

     (i) SEQUENCE CHARACTERISTICS:
           (A) LENGTH: 43 base pairs
           (B) TYPE: nucleic acid
           (C) STRANDEDNESS: single
           (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: other nucleic acid
           (A) DESCRIPTION: /desc = "DNA"

    (ix) FEATURE:
           (A) NAME/KEY: misc_difference
           (B) LOCATION: replace(1, "")
           (D) OTHER INFORMATION: /note= "The residue at this
  position is a spacer containing a fluorescein label"

    (ix) FEATURE:
           (A) NAME/KEY: modified_base
           (B) LOCATION: 23
           (D) OTHER INFORMATION: /note= "The residue at this
  position indicates 2' o-methyl sugar"
```

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(43, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3' amine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

NGTCTGAGAT GAAAGTGCTC CCGCACCCAC CCAAGGCACA GCN 43

(2) INFORMATION FOR SEQ ID NO:184:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1..17
(D) OTHER INFORMATION: /note= "The residues at these positions are 2' o-methyl sugars"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(18, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3' amine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

GCTGTGCCTT GGGTGGGN 18

(2) INFORMATION FOR SEQ ID NO:185:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(29, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3' primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

AACGAGGCGC ACCCACCCAA GGCACAGCN 29

(2) INFORMATION FOR SEQ ID NO:186:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

GCTGTGCCTT GGGTGGGTGC G 21

(2) INFORMATION FOR SEQ ID NO:187:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(21, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3' amine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

GCTGTGCCTT GGGTGGGTGC N 21

(2) INFORMATION FOR SEQ ID NO:188:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 15..20
(D) OTHER INFORMATION: /note= "The residues at these positions are 2' o-methyl sugars"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(21, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3' amine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

GCTGTGCCTT GGGTGGGTGC N 21

(2) INFORMATION FOR SEQ ID NO:189:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 1..20
(D) OTHER INFORMATION: /note= "The residues at these positions are 2' o-methyl sugars"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(21, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a 3' amine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:189:

GCTGTGCCTT GGGTGGGTGC N 21

(2) INFORMATION FOR SEQ ID NO:190:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "DNA"

(ix) FEATURE:
(A) NAME/KEY: misc_difference
(B) LOCATION: replace(1, "")
(D) OTHER INFORMATION: /note= "The residue at this position is a spacer containing a fluorescein label"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

NTCTTCGCAC ATTTCATCTC AGACGGA 27

CLAIMS

1. A method of detecting the presence of a target nucleic acid molecule comprising:

a) providing:

i) a cleavage means,

ii) a source of a first target nucleic acid, said first target nucleic acid having a first region, a second region and a third region, wherein said first region is located adjacent to and downstream from said second region and wherein said second region is located adjacent to and downstream from said third region;

iii) a first oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said first oligonucleotide contains a sequence complementary to said second region of said first target nucleic acid and wherein said 3' portion of said first oligonucleotide contains a sequence complementary to said third region of said first target nucleic acid;

iv) a second oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said second oligonucleotide contains a sequence complementary to said first region of said first target nucleic acid and wherein said 3' portion of said second oligonucleotide contains a sequence complementary to said second region of said first target nucleic acid;

v) a source of a second target nucleic acid, said second target nucleic acid having a first region, a second region and a third region, wherein said first region is located adjacent to and downstream from said second region and wherein said second region is located adjacent to and downstream from said third region;

vi) a third oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said third oligonucleotide contains a sequence complementary to said second region of said second target nucleic acid and wherein said 3' portion of said third oligonucleotide contains a sequence complementary to said third region of said second target nucleic acid;

b) generating a first cleavage structure wherein at least said 3' portion of said first oligonucleotide is annealed to said first target nucleic acid and wherein at least said 5' portion of said second oligonucleotide is annealed to said first target nucleic acid and wherein cleavage of said first cleavage structure occurs via said cleavage means thereby cleaving said first oligonucleotide to generate a fourth oligonucleotide, said fourth oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said fourth oligonucleotide contains a sequence complementary to said first region of said second target nucleic acid and wherein said 3' portion of said fourth oligonucleotide contains a sequence complementary to said second region of said second target nucleic acid;

c) generating a second cleavage structure under conditions wherein said at least said 3' portion of said third oligonucleotide is annealed to said second target nucleic acid and wherein at least said 5' portion of said fourth oligonucleotide is annealed to said second target nucleic acid oligonucleotide and wherein cleavage of said second cleavage structure occurs to generate a fifth oligonucleotide, said fifth oligonucleotide having a 3'-hydroxyl group; and

d) detecting said fifth oligonucleotide.

2. The method of Claim 1, wherein said first oligonucleotide has a length between eleven and fifteen nucleotides.

3. The method of Claim 1, wherein said cleavage means is a structure-specific nuclease.

4. The method of Claim 3, wherein said structure-specific nuclease is a thermostable structure-specific nuclease.

5. The method of Claim 4, wherein said thermostable structure-specific nuclease is a *Pyrococcus woesii* FEN-1 endonuclease.

6. The method of Claim 1, wherein one or more of said first, second, and said third oligonucleotides contain a dideoxynucleotide at the 3' terminus.

7. The method of Claim 1, wherein said detecting said fifth oligonucleotide comprises:

a) incubating said fifth oligonucleotide with a template-independent polymerase and at least one labelled nucleoside triphosphate under conditions such that at least one labelled nucleotide is added to the 3'-hydroxyl group of said fifth olignucleotide to generate a labelled fifth oligonucleotide; and

b) detecting the presence of said labelled fifth oligonucleotide.

8. The method of Claim 7, wherein said template-independent polymerase is selected from the group consisting of terminal deoxynucleotidyl transferase and poly A polymerase.

9. The method of Claim 8, wherein said third oligonucleotide contains a 5' end label, said 5' end label being a different label than the label present upon said labelled nucleoside triphosphate.

10. The method of Claim 1, wherein said detecting said fifth oligonucleotide comprises:

a) incubating said fifth oligonucleotide with a template-independent polymerase and at least one nucleoside triphosphate under conditions such that at least one nucleotide is added to the 3'-hydroxyl group of said fifth oligonucleotide to generate a tailed oligonucleotide; and

b) detecting the presence of said tailed fifth oligonucleotide.

11. The method of Claim 10, wherein said template-independent polymerase is selected from the group consisting of terminal deoxynucleotidyl transferase and poly A polymerase.

12. The method of Claim 11, wherein said third oligonucleotide contains a 5' end label.

13. The method of Claim 1, wherein said detecting said fifth oligonucleotide comprises:

a) providing:

i) said fifth oligonucleotide;

ii) a composition comprising two single-stranded nucleic acids annealed so as to define a single-stranded portion of a protein binding region;

iii) a nucleic acid producing protein;

b) exposing said fifth oligonucleotide to said single-stranded portion of said protein binding region under conditions such that said nucleic acid producing protein binds to said protein binding region and produces nucleic acid.

14. The method of Claim 13, wherein said single-stranded portion of said protein binding region comprises:

a) a first single continuous strand of nucleic acid comprising a sequence defining the template strand of an RNA polymerase binding region; and

b) a second single continuous strand of nucleic acid having a 5' and a 3' end, said second nucleic acid comprising a region complementary to a portion of said first nucleic acid, wherein said second nucleic acid is annealed to said first nucleic acid so as to define said single-stranded portion of said protein binding region.

15. The method of Claim 14, wherein said protein binding region is a template-dependent RNA polymerase binding region.

16. The method of Claim 15, wherein said template-dependent RNA polymerase binding region is the T7 RNA polymerase binding region.

17. The method of Claim 1, wherein said detecting said fifth oligonucleotide comprises:

a) providing:

i) said fifth oligonucleotide;

ii) a single continuous strand of nucleic acid comprising a sequence defining a single strand of an RNA polymerase binding region;

iii) a template-dependent DNA polymerase;

iv) a template-dependent RNA polymerase;

b) exposing said fifth oligonucleotide to said RNA polymerase binding region under conditions such that said fifth oligonucleotide binds to a portion of said single strand of said RNA polymerase binding region;

c) exposing said bound fifth oligonucleotide to said template-dependent DNA polymerase under conditions such that a double-stranded RNA polymerase binding region is produced; and

d) exposing said double-stranded RNA polymerase binding region to said template-dependent RNA polymerase under conditions such that RNA transcripts are produced.

18. The method of Claim 17 further comprising detecting said RNA transcripts.

19. The method of Claim 17, wherein said template-dependent RNA polymerase is the T7 RNA polymerase.

20. A method of detecting the presence of a target nucleic acid molecule comprising:

a) providing:

i) a cleavage means,

ii) a source of a first target nucleic acid, said first target nucleic acid having a first region, a second region, a third region and a fourth region, wherein said first region is located adjacent to and downstream from said second region, said second region is located adjacent to and downstream from said third region and said third region is located adjacent to and downstream from said fourth region;

iii) a first oligonucleotide complementary to said fourth region of said first target nucleic acid;

iv) a second oligonucleotide having a 5' portion and a 3' portion wherein said 5' portion of said second oligonucleotide contains a sequence complementary to said second region of said first target nucleic acid and wherein said 3' portion of said second oligonucleotide contains a sequence complementary to said third region of said first target nucleic acid;

iv) a third oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said third oligonucleotide contains a sequence complementary to said first region of said first target nucleic acid and wherein said 3' portion of said third oligonucleotide contains a sequence complementary to said second region of said first target nucleic acid;

v) a source of a second target nucleic acid, said second target nucleic acid having a first region, a second region and a third region, wherein said first region is located adjacent to and downstream from said second region and wherein said second region is located adjacent to and downstream from said third region;

vi) a fourth oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said fourth oligonucleotide contains a sequence complementary to said second region of said second target nucleic acid and wherein said 3' portion of said fourth oligonucleotide contains a sequence complementary to said third region of said second target nucleic acid;

b) generating a first cleavage structure wherein said first oligonucleotide is annealed to said fourth region of said first target nucleic acid and wherein at least said 3' portion of said second oligonucleotide is annealed to said first target nucleic acid and wherein at least said 5' portion of said third oligonucleotide is annealed to said first target nucleic acid and wherein cleavage of said first cleavage structure occurs thereby cleaving said second oligonucleotide to generate a fifth oligonucleotide, said fifth oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said fifth oligonucleotide contains a sequence complementary to said first region of said second target nucleic acid and wherein said 3' portion of said fifth oligonucleotide contains a sequence complementary to said second region of said second target nucleic acid;

c) generating a second cleavage structure under conditions wherein said at least said 3' portion of said fourth oligonucleotide is annealed to said second target nucleic acid and wherein at least said 5' portion of said fifth oligonucleotide is annealed to said second target nucleic acid and wherein cleavage of said second cleavage structure occurs to generate a sixth oligonucleotide, said sixth oligonucleotide having a 3'-hydroxyl group; and

d) detecting said sixth oligonucleotide.

21. A method of detecting the presence of a target nucleic acid molecule comprising:

a) providing:

i) a cleavage means,

ii) a source of a target nucleic acid, said target nucleic acid having a first region, a second region and a third region, wherein said first region is located adjacent to and downstream from said second region and wherein said second region is located adjacent to and downstream from said third region;

iii) a first oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said first oligonucleotide contains a sequence complementary to said second region of said target nucleic acid and wherein said 5' portion of said first oligonuceotide contains a region of self-complementarity and wherein said 3' portion of said first oligonucleotide contains a sequence complementary to said third region of said target nucleic acid;

iv) a second oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said second oligonucleotide contains a sequence complementary to said first region of said target nucleic acid and wherein said 3' portion of said second oligonucleotide contains a sequence complementary to said second region of said target nucleic acid;

v) a third oligonucleotide having a 5' and a 3' portion wherein said 3' portion of said third oligonucleotide contains a sequence complementary to said 5' portion of said first oligonucleotide;

b) generating a first cleavage structure wherein at least said 3' portion of said first oligonucleotide is annealed to said target nucleic acid and wherein at least said 5' portion of said second oligonucleotide is annealed to said target nucleic acid and wherein cleavage of said first cleavage structure occurs thereby cleaving said first oligonucleotide to generate a fourth oligonucleotide, said fourth oligonucleotide having a first region, a second region and a third region, wherein said first region is located adjacent to and upstream of said second region and wherein said second region is located adjacent to and upstream of said third region and wherein said third region of said fourth oligonucleotide contains a a region of self-complementarity;

c) generating a second cleavage structure under conditions wherein said at least said 3' portion of said third oligonucleotide is annealed to said first region of

said fourth oligonucleotide and wherein at least said 5' portion of said fourth oligonucleotide is annealed to said second region of said third olignucleotide and wherein said third region of said fourth oligonucleotide forms a hairpin structure and wherein cleavage of said second cleavage structure occurs to generate a fifth oligonuceotide, said fifth oligonucleotide having a 3'-hydroxyl group; and

d) detecting said fifth oligonucleotide.

22. A method of detecting the presence of human cytomegalovirus nucleic acid in a sample comprising:

a) providing:

i) a cleavage means,

ii) a sample suspected of containing human cytomegalovirus target nucleic acid, said target nucleic acid having a first region, a second region and a third region, wherein said first region is located adjacent to and downstream from said second region and wherein said second region is located adjacent to and downstream from said third region;

iii) a first oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said first oligonucleotide contains a sequence complementary to said second region of said target nucleic acid and wherein said 3' portion of said first oligonucleotide contains a sequence complementary to said third region of said target nucleic acid;

iv) a second oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said second oligonucleotide contains a sequence complementary to said first region of said target nucleic acid and wherein said 3' portion of said second oligonucleotide contains a sequence complementary to said second region of said target nucleic acid;

b) generating a cleavage structure wherein at least said 3' portion of said first oligonucleotide is annealed to said target nucleic acid and wherein at least said 5' portion of said second oligonucleotide is annealed to said target nucleic acid and wherein cleavage of said cleavage structure occurs via said cleavage means to generate non-target cleavage products, each non-target cleavage product having a 3' hydroxyl group; and

c) detecting said non-target cleavage products and thereby detecting the presence of human cytomegalovirus nucleic acid in said sample.

23. The method of Claim 22, wherein said first oligonucleotide has a length between eleven and fifteen nucleotides.

24. The method of Claim 22, wherein said cleavage means is a thermostable structure-specific nuclease.

25. The method of Claim 24, wherein said nuclease is a *Pyrococcus woesii* FEN-1 endonuclease.

26. The method of Claim 22, wherein one or more of said first and said second oligonucleotides contain a dideoxynucleotide at the 3' terminus.

27. The method of Claim 22, wherein said detecting said non-target cleavage products comprises:

a) incubating said non-target cleavage products with a template-independent polymerase and at least one labelled nucleoside triphosphate under conditions such that at least one labelled nucleotide is added to the 3'-hydroxyl group of said non-target cleavage products to generate labelled non-target cleavage products; and

b) detecting the presence of said labelled non-target cleavage products.

28. The method of Claim 22, wherein said detecting said non-target cleavage products comprises:

a) incubating said non-target cleavage products with a template-independent polymerase and at least one nucleoside triphosphate under conditions such that at least one nucleotide is added to the 3'-hydroxyl group of said non-target cleavage products to generate tailed non-target cleavage products; and

b) detecting the presence of said tailed non-target cleavage products.

29. The method of Claim 22, wherein said detecting said non-target cleavage products comprises:

a) providing:

i) said non-target cleavage products;

ii) a composition comprising two single-stranded nucleic acids annealed so as to define a single-stranded portion of a protein binding region;

iii) a nucleic acid producing protein;

b) exposing said non-target cleavage products to said single-stranded portion of said protein binding region under conditions such that said nucleic acid producing protein binds to said protein binding region and produces nucleic acid.

30. The method of Claim 29, wherein said single-stranded portion of said protein binding region comprises:

a) a first single continuous strand of nucleic acid comprising a sequence defining the template strand of an RNA polymerase binding region; and

b) a second single continuous strand of nucleic acid having a 5' and a 3' end, said second nucleic acid comprising a region complementary to a portion of said first nucleic acid, wherein said second nucleic acid is annealed to said first nucleic acid so as to define said single-stranded portion of said protein binding region.

31. The method of Claim 29, wherein said protein binding region is a template-dependent RNA polymerase binding region.

32. The method of Claim 31, wherein said template-dependent RNA polymerase binding region is the T7 RNA polymerase binding region.

33. The method of Claim 22, wherein said detecting said non-target cleavage products comprises:

a) providing:

i) said non-target cleavage products;

ii) a single continuous strand of nucleic acid comprising a sequence defining a single strand of an RNA polymerase binding region;

iii) a template-dependent DNA polymerase;

iv) a template-dependent RNA polymerase;

b) exposing said non-target cleavage products to said RNA polymerase binding region under conditions such that said non-target cleavage product binds to a portion of said single strand of said RNA polymerase binding region;

c) exposing said bound non-target cleavage products to said template-dependent DNA polymerase under conditions such that a double-stranded RNA polymerase binding region is produced;

d) exposing said double-stranded RNA polymerase binding region to said template-dependent RNA polymerase under conditions such that RNA transcripts are produced.

34. A method of detecting the presence of human cytomegalovirus nucleic acid in a sample comprising:

a) providing:

i) a cleavage means,

ii) a sample suspected of containing human cytomegalovirus target nucleic acid, said target nucleic acid having a first region, a second region, a third region and a fourth region, wherein said first region is located adjacent to and downstream from said second region, said second region is located adjacent to and downstream from said third region and said third region is located adjacent to and downstream from said fourth region;

iii) a first oligonucleotide complementary to said fourth region of said target nucleic acid;

iv) a second oligonucleotide having a 5' portion and a 3' portion wherein said 5' portion of said second oligonucleotide contains a sequence complementary to said second region of said target nucleic acid and wherein said 3' portion of said first oligonucleotide contains a sequence complementary to said third region of said target nucleic acid;

v) a third oligonucleotide having a 5' portion and a 3' portion wherein said 5' portion of said third oligonucleotide contains a sequence complementary to said first region of said target nucleic acid and wherein said 3' portion of said third oligonucleotide contains a sequence complementary to said second region of said target nucleic acid;

b) generating a cleavage structure wherein said first oligonucleotide is annealed to said fourth region of said target nucleic acid and wherein at least said 3' portion of said second oligonucleotide is annealed to said target nucleic acid and wherein at least said 5' portion of said third oligonucleotide is annealed to said target nucleic acid and wherein cleavage of said cleavage structure occurs via said cleavage means to generate non-target cleavage products, each non-target cleavage product having a 3' hydroxyl group; and detecting said non-target cleavage products and thereby detecting the presence of human cytomegalovirus nucleic acid in said sample.

35. A method of detecting the presence of a target nucleic acid molecule comprising:

a) providing:

i) a cleavage means,

ii) a source of a first target nucleic acid, said first target nucleic acid having a first region, a second region and a third region, wherein said first region is located adjacent to and downstream from said second region and wherein said second region is located adjacent to and downstream from said third region;

iii) first and second oligonucleotides having 3' and 5' portions, wherein said 3' portion of said first oligonucleotide contains a sequence complementary to said third region of said target nucleic acid and wherein said 5' portion of said first oligonucleotide and said 3' portion of said second oligonucleotide each contain sequence full complementary to said second region of said target nucleic acid, and wherein said 5' portion of said second oligonucleotide contains sequence complementary to said first region of said target nucleic acid;

iv) a source of a second target nucleic acid, said second target nucleic acid having a first region, a second region and a third region, wherein said first region is located adjacent to and downstream from said second region and wherein said second region is located adjacent to and downstream from said third region;

v) a third oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said third oligonucleotide contains a sequence complementary to said second region of said second target nucleic acid and wherein said 3' portion of said third oligonucleotide contains a sequence complementary to said third region of said second target nucleic acid;

b) generating a first cleavage structure wherein at least said 3' portion of said first oligonucleotide is annealed to said first target nucleic acid and wherein at least said 5' portion of said second oligonucleotide is annealed to said first target nucleic acid and wherein cleavage of said first cleavage structure occurs via said cleavage means thereby cleaving said first oligonucleotide to generate a fourth oligonucleotide, said fourth oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said fourth oligonucleotide contains a sequence complementary to said first region of said second target nucleic acid and wherein said 3' portion of said fourth oligonucleotide contains a sequence complementary to said second region of said second target nucleic acid;

c) generating a second cleavage structure under conditions wherein said at least said 3' portion of said third oligonucleotide is annealed to said second target nucleic acid and wherein at least said 5' portion of said fourth oligonucleotide is annealed to said second target nucleic acid oligonucleotide and wherein cleavage of said second cleavage structure occurs to generate a fifth oligonucleotide, said fifth oligonucleotide having a 3'-hydroxyl group; and

d) detecting said fifth oligonucleotide.

36. The method of Claim 35, wherein said first oligonucleotide has a length between eleven and fifteen nucleotides.

37. The method of Claim 35, wherein said cleavage means is a structure-specific nuclease.

38. The method of Claim 37, wherein said structure-specific nuclease is a thermostable structure-specific nuclease.

39. The method of Claim 38, wherein said thermostable structure-specific nuclease is an *Afu* FEN-1 endonuclease.

40. The method of Claim 35, wherein one or more of said first, second, and said third oligonucleotides contain a dideoxynucleotide at the 3' terminus.

41. The method of Claim 35, further comprising providing an Arrestor™, wherein said Arrestor™ reduces interaction between said first oligonucleotide and said second target.

42. The method of Claim 35, wherein said detecting said fifth oligonucleotide comprises:

a) incubating said fifth oligonucleotide and at least one labelled nucleoside triphosphate under conditions such that at least one labelled nucleotide is added to the 3'-hydroxyl group of said fifth olignucleotide to generate a labelled fifth oligonucleotide; and

b) detecting the presence of said labelled fifth oligonucleotide.

43. The method of Claim 42, wherein said step a) further comprises incubating a polymerase with said fifth oligonucleotide and at least one labelled nucleoside triphosphate.

44. The method of Claim 43, wherein said polymerase is a template-dependent polymerase.

45. The method of Claim 44, wherein said template-dependent polymerase is selected from the group consisting of terminal deoxynucleotidyl transferase and poly A polymerase.

46. The method of Claim 35, wherein said third oligonucleotide contains a 5' end label, said 5' end label being a different label than the label present upon said labelled nucleoside triphosphate.

47. The method of Claim 35, wherein said detecting of said fifth oligonucleotide comprises:

a) incubating said fifth oligonucleotide with a polymerase and at least one nucleoside triphosphate under conditions such that at least one nucleotide is added to the 3'-hydroxyl group of said fifth oligonucleotide to generate a tailed oligonucleotide; and

b) detecting the presence of said tailed fifth oligonucleotide.

48. The method of Claim 47, wherein said polymerase is a template-dependent polymerase.

49. The method of Claim 48, wherein said template-dependent polymerase is selected from the group consisting of terminal deoxynucleotidyl transferase and poly A polymerase.

50. The method of Claim 49, wherein said third oligonucleotide contains a 5' end label.

51. The method of Claim 47, wherein said detecting said fifth oligonucleotide comprises:

a) providing:

i) said fifth oligonucleotide;

ii) a single continuous strand of nucleic acid comprising a sequence definining a single strand of an RNA polymerase binding region;

iii) a template-dependent DNA polymerase;

iv) a template-dependent RNA polymerase;

b) exposing said fifth oligonucleotide to said RNA polymerase binding region under conditions such that said fifth oligonucleotide binds to a portion of said single strand of said RNA polymerase binding region to produce a bound fifth oligonucleotide;

c) exposing said bound fifth oligonucleotide to said template-dependent DNA polymerase under conditions such that a double-stranded RNA polymerase binding region is produced; and

d) exposing said double-stranded RNA polymerase binding region to said template-dependent RNA polymerase under conditions such that RNA transcripts are produced.

52. The method of Claim 51, wherein said single-stranded portion of said protein binding region comprises:

a) a first single continuous strand of nucleic acid comprising a sequence defining the template strand of an RNA polymerase binding region; and

b) a second single continuous strand of nucleic acid having a 5' and a 3' end, said second nucleic acid comprising a region complementary to a portion of said first nucleic acid, wherein said second nucleic acid is annealed to said first nucleic acid so as to define said single-stranded portion of said protein binding region.

53. The method of Claim 52, wherein said protein binding region is a template-dependent RNA polymerase binding region.

54. The method of Claim 53, wherein said template-dependent RNA polymerase binding region is the T7 RNA polymerase binding region.

55. The method of Claim 35, wherein said detecting said fifth oligonucleotide comprises:

a) providing:

i) said fifth oligonucleotide;

ii) a single continuous strand of nucleic acid comprising a sequence defining a single strand of an RNA polymerase binding region;

iii) a template-dependent DNA polymerase;

iv) a template-dependent RNA polymerase;

b) exposing said fifth oligonucleotide to said RNA polymerase binding region under conditions such that said fifth oligonucleotide binds to a portion of said single strand of said RNA polymerase binding region;

c) exposing said bound fifth oligonucleotide to said template-dependent DNA polymerase under conditions such that a double-stranded RNA polymerase binding region is produced; and

d) exposing said double-stranded RNA polymerase binding region to said template-dependent RNA polymerase under conditions such that RNA transcripts are produced.

56. The method of Claim 55, further comprising detecting said RNA transcripts.

57. The method of Claim 55, wherein said template-dependent RNA polymerase is the T7 RNA polymerase.

58. A method of detecting the presence of a target nucleic acid molecule comprising:

a) providing:

i) a cleavage means,

ii) a source of a first target nucleic acid, said first target nucleic acid having a first region, a second region, a third region and a fourth region, wherein said first region is located adjacent to and downstream from said second region, said second region is located adjacent to and downstream from said third region and said third region is located adjacent to and downstream from said fourth region;

iii) a first oligonucleotide complementary to said fourth region of said first target nucleic acid;

iv) second and third oligonucleotides having 3' and 5' portions, wherein said 3' portion of said second oligonucleotide contains a sequence complementary to said third region of said target nucleic acid and wherein said 5' portion of said second oligonucleotide and said 3' portion of said third oligonucleotide each contain sequence completely complementary to said second region of said target nucleic acid, and wherein said 5' portion of said third oligonucleotide contains sequence complementary to said first region of said target nucleic acid;

v) a source of a second target nucleic acid, said second target nucleic acid having a first region, a second region and a third region, wherein said first region is located adjacent to and downstream from said second region

and wherein said second region is located adjacent to and downstream from said third region;

vi) a fourth oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said fourth oligonucleotide contains a sequence complementary to said second region of said second target nucleic acid and wherein said 3' portion of said fourth oligonucleotide contains a sequence complementary to said third region of said second target nucleic acid;

b) generating a first cleavage structure wherein said first oligonucleotide is annealed to said fourth region of said first target nucleic acid and wherein at least said 3' portion of said second oligonucleotide is annealed to said first target nucleic acid and wherein at least said 5' portion of said third oligonucleotide is annealed to said first target nucleic acid and wherein cleavage of said first cleavage structure occurs thereby cleaving said second oligonucleotide to generate a fifth oligonucleotide, said fifth oligonucleotide having a 5' and a 3' portion wherein said 5' portion of said fifth oligonucleotide contains a sequence complementary to said first region of said second target nucleic acid and wherein said 3' portion of said fifth oligonucleotide contains a sequence complementary to said second region of said second target nucleic acid;

c) generating a second cleavage structure under conditions wherein said at least said 3' portion of said fourth oligonucleotide is annealed to said second target nucleic acid and wherein at least said 5' portion of said fifth oligonucleotide is annealed to said second target nucleic acid and wherein cleavage of said second cleavage structure occurs to generate a sixth oligonucleotide, said sixth oligonucleotide having a 3'-hydroxyl group; and

d) detecting said sixth oligonucleotide.

59. The method of Claim 58, wherein said first oligonucleotide has a length between eleven and fifteen nucleotides.

60. The method of Claim 58, wherein said cleavage means is a structure-specific nuclease.

61. The method of Claim 60, wherein said structure-specific nuclease is a thermostable structure-specific nuclease.

62. The method of Claim 61, wherein said thermostable structure-specific nuclease is an *Afu* FEN-1 endonuclease.

63. The method of Claim 58, wherein one or more of said first, second, and said third oligonucleotides contain a dideoxynucleotide at the 3' terminus.

64. The method of Claim 58, further comprising providing an Arrestor™, wherein said Arrestor™ reduces interaction between said first oligonucleotide and said second target.

65. The method of Claim 64, wherein said detecting said sixth oligonucleotide comprises:

a) incubating said sixth oligonucleotide and at least one labelled nucleoside triphosphate under conditions such that at least one labelled nucleotide is added to the 3'-hydroxyl group of said sixth olignucleotide to generate a labelled sixth oligonucleotide; and

b) detecting the presence of said labelled sixth oligonucleotide.

66. The method of Claim 65, wherein said step a) further comprises incubating a polymerase with said sixth oligonucleotide and at least one labelled nucleoside triphosphate.

67. The method of Claim 66, wherein said polymerase is a template-dependent polymerase.

68. The method of Claim 67, wherein said template-dependent polymerase is selected from the group consisting of terminal deoxynucleotidyl transferase and poly A polymerase.

FIGURE 1A

```
MAJORITY  ATGXXGGCGATGCTTCCCCTCTTTGAGCCCAAAGGCCGGGTCCTCCTGGTGGACGGCCACCACCTGGCCT

DNAPTAQ   ...AG..G......G.................G.....................................  70
DNAPTFL   .....................................C..G.............................  67
DNAPTTH   ...GA............G........A...........................................  70

MAJORITY  ACCGCACCXTCTTCGCCCTGAAGGGCCTCACCACCAGCCGGGCCGAACCGGTGCAGGCGGTCTACGCCTT

DNAPTAQ   ...........CA..............................G..G.......................  140
DNAPTFL   .............T.....C....................C........C..T.................  137
DNAPTTH   ..................................G...................................  140

MAJORITY  CGCCAAGAGCCTCCTCAAGGCCCTGAAGGAGGACGGGGACXXGGCGGTGXTCGTGGTCTTTGACGCCAAG

DNAPTAQ   ........................C........................A....................  207
DNAPTFL   ......A.................................GT..T.........................  204
DNAPTTH   .....................................T..AA...C..CT....................  210

MAJORITY  GCCCCCTCCTTCCGCCACGAGGCCTACGAGGCCTACAAGGCGGGCCGGGCCCCCACCCCGGAGGACTTTC

DNAPTAQ   ............................G..GG.......................G.............  277
DNAPTFL   ......................................................................  274
DNAPTTH   ............................................GA.......G.....C........C.  280

MAJORITY  GCCGCCAGCTCGCCCTCATCAAGGAGCTGGTGGACCTCCTGGGCCTTGCGCGCCTCGAGGTCCCCGGCTA

DNAPTAQ   .......A......................................G................G......  347
DNAPTFL   ..........G...............T.............A..C....T..G...G....T........T  344
DNAPTTH   ............................................T..A.C....................  350
```

FIGURE 1B

```
[illegible]   CCAGCCGGACGACGTXCTGGCCACCCTGGCCAAGAAGGCGGAAAAGGAGGGGTACGACGTGCGCATCCTC

[illegible]   ..............C.......G...............................C........C.......... 417
[illegible]   T.............G.................CG....................................... 414
[illegible]   ..............T..C....................................................... 420

[illegible]   ACCCCCGACGGCGACCTCTACCAGCTCCTTTCCGACCGCATCGCCGTCCTCCACCCCGAGGGGTACCTCA

[illegible]   .......AAA.....T......................CA.............................. 487
[illegible]   ..T......................C..G........A.........T..........G. 484
[illegible]   ....................A...G.C.........G.......................CC...... 490

[illegible]   TCACCCCGGCGTGCCTTTGGGAGAAGTACGGCCTGAGGCCGGAGCAGTGGGTGGACTACCGGGCCCTGGC

[illegible]   .....C..........A..................C..C......CC...............A. 557
[illegible]   .............AC...............C.C................................ 554
[illegible]   .A.....................C......................T...C....C.T 560

[illegible]   GGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCCXGAAGCTCCTCXAG

[illegible]   C    ..GAG..........T...........................G..GAG......T..GG.. 627
[illegible]   ......G..T...A..........G...........................A..G....A..CGC 624
[illegible]   ...................................................TC.........A.. 630

[illegible]   GAGTGGGGGAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTGAAGCCCCGC--CXTCCGGGAGAAGA

[illegible]   ...............GC......................C...........A............. 694
[illegible]   .........................T..C..C........A..........T.....T.G.........C 691
[illegible]   .....A..................................A.....A.AAAA.G............ 700
```

FIGURE 1C

```
TCCAGGCCCACATGGAXGACCTGAXGCTCTCCTGGGAGCTXTCCCAGGTGCGCACCCACCTGCCCCTGCA

   T   ...........C..T....A............C..GG..A....................  ....  784
      .GGG......G.C....GCC..T...C..A....T..........A...T...............  761
 .A   ......C....A.....C.G.......T......C....G...............C........  770

CGTGGACTTCGCCAAGXGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGCCTGGAGTTT

     .........AA...................A.......................T......  834
  .......GG.G.C.C..CACA...A...T......T..GC....T...T.....C..T........  831
     .C......C..G..............................C.................C  840

GGCACCCTCCTCCACGAGTTCGGCCTCCTGGAGGGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGC

. ..  ......................T.....AA....................................  904
 A..............................G..G.....GGCA.....................T.  901
  .  ..........................C....GCCC...........................  910

CGGAAGGGGCCTTCGTGGGCTTTGTCCTTTCCCGCCCCCAGCCCATGTGGCCCGAGCTTCTGGCCCTGGC

   .  ...................G.........AAG.................T..............  974
    . .T..TT........TC.T.......T............................................  971
      ... .........C.....C........................G......AAA........  980

CGCCGCCAGGGAGGGCCGGGTCCACCGGGCACCAGACCCCTTTAXGGGCCTXAGGGACCTXAAGGACGTG

       .....G..................C..C..G..T.A..AA.C...C........G.......C.  1044
T  GG..GT.......G..CC....T.......A.......C...G......G........T....G....  1041
    TC......C........G.........G..........GGC...G..A.A......C........C  1050
```

FIGURE 1D

```
CGGGGXCTCCTCGCCAAGGACCTGGCCGTTTTGGCCCTGAGGGAGGGCCTXGACCTCXTGCCCGGGGACG
     C..T........A......AG....C..............A.....T.G....CC......C....  1114
     AA....G...............G..............C..........G......T.C..A.A.....  1111
    .C.................C.....C......TC..........G..A......G............  1120

ACCCCATGCTCCTCGCCTACCTCCTGGACCCCTCCAACACCACCCCCGAGGGGTGGCCCCGGCGCTACGG
   .............................T............................................  1184
   .......G.......T.......................T................T.....  1181
   ...........................................................G...........  1190

GGGGCAGTGCACGGAGGAXGCGGGGGAGCGGGGCCCTCCTXTCCGAGAGGCTCTTCCXGAACCTXXXGGAG
C..................G...............GC...T................GCC.....GTG..G.  1254
...................T.........A..........GG.....C.C.......A..C...AAA....  1251
...................C..C.CCC.C...........C..G.........CAT.G......CCTTA..  1260

GGCCTTGAGCGGGACGAGAGGCTCCTTTGGCTTTACCAGGAGCTGGAGAAGCCCCTTTCCCGGGTCCTGG
............................G............G..........GCT.......  1324
A  G .........................G.................G..........GT...  1321
  A.   A..A..AC.C..C............C.....C.........A........C.............  1330
C  ..........A...........................................................

GCCACATGGAGGCCACGGGGGTXCGGCTGGACGTGGCCTACCTCCAGGCCCTXTCCCTGGAGGTGGCGGA
.....................G..C...............T...AG....T.G..............C..  1394
.....................C...................................C.............A..C  1391
CG.....................C..........................T.........T.........C.T.....  1400
..............C.....A...........................................................
```

FIGURE 1E

```
MAJORITY  GGAGATCCCCCCCCTCGAGGAGGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCGGGAC
INVPTAO   .......GC...........CC............................................... 1464
INVPTFL   ...G.G....AG..G.............................................C...    1461
INVPTTH   ..................................T....G...........................   1470

MAJORITY  CACCTGGAAAGGGTGCTCTTTGACGAGCTXGGGCTTCCCGCCATCGGCAAGACCGAGAAGACXGGCAAGC
INVPTAO   ..........C..............A..........................C.......         1534
INVPTFL   ....CC...................G..C..G..T.....................G..G..A.     1531
INVPTTH   .........................TA...........T.G..G......C.A.....A.......   1540

MAJORITY  GCTCCACCAGCGCCGCCGTGCTGGAGGCCCTXCGXGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTA
INVPTAO   ...............C...........C..C....................................  1604
INVPTFL   ..........T.................G..A.....................CCCC...........  1601
INVPTTH   ..........G...............A..G..................................C...C. 1610

MAJORITY  CCGGGAGCTCACCAAGCTCAAGAACACCTACATXGACCCCCTGCCXGXCCTCGTCCACCCCAGGACCGGC
INVPTAO   ..................G....G.........T......T....G.A....A................ 1674
INVPTFL   ...........................A...........C.C...G..........A...C...      1671
INVPTTH   ..........................G.G........C..AAG............G.........     1680

MAJORITY  CGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGCCTTAGTAGCTCCGACCCCAACCTGC
INVPTAO   ...................................A...........T........C.           1744
INVPTFL   G......................C...............TCC...........................  1741
INVPTTH   ..............................G..........................             1750
```

FIGURE 1F

```
MAJORITY   AGAACATCCCCGTCCGCACCCCXCTGGGCCAGAGGATCCGCCGGGCCTTCGTGGCCGAGGAGGGXTGGGT

DNAPTAQ    . ....................G..T..G.....................A.C...........G...C.  1814
DNAPTFL     .. ........G........T.........C.C........A...................C.....  1811
DNAPTTH    . ....................CT...............................C...T....C  1820

MAJORITY   GTTGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGGGTCCTGGCCCACCTCTCCGGGGACGAGAACCTG

DNAPTAQ    A.................................A....G..................C.........  ...  1894
DNAPTFL     C    ....T.T.......C.........T.....T...................................  1881
DNAPTTH    . ...................................C.....C..................A......  1890

MAJORITY   ATCCGGGTCTTCCAGGAGGGGAGGGACATCCACACCCAGACCGCCAGCTGGATGTTCGGCGTCCCCCCGG

DNAPTAQ    . .....................C.............GG..............................G..  1954
DNAPTFL    . ...........T.........................................TT....C.  1951
DNAPTTH     ..A..................A...................A........................  1960

MAJORITY   AGGCCGTGGACCCCCTGATGCCCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTCCGC

DNAPTAQ     .... ............................................................C..  2024
DNAPTFL    .A.GG. A.....T..................................G.................  2021
DNAPTTH          .. ........ ...................GG.C.......C................ ...  2030

MAJORITY   CCACCGCCTCTCCCAGGAGCTTGCCATCCCCTACGAGGAGGCGGTGGCCTTCATTGAGCGCTACTTCCAG

DNAPTAQ     .  ................A........T............CCA.......... .........T .  2094
DNAPTFL     .............GG.......T.........................................  2091
DNAPTTH    .  TA..G.,........................................T..A..............A  2100
```

6/133

FIGURE 1G

```
MAJORITY   ACCTTCCCCAAGGTGCGGGCCTGCATTGAGAAGACCCTGGAGGAGGCCAGGAGGCGGGGCTACCTGGAGA

GBVPTAO    .. .....................................................................  2164
GBVPTRL       A.......................GG.......C.........C.CC..........T.......  2161
GBVPTTH    ...  ....................A..A..................G....A......C........A.  2170

MAJORITY   CCCTCTTCGGCCGCCGGCGCTACGTGCCCGACCTCAACGCCCCGGTGAAGAGCGTGCGGGAGGCGGCGGA

GBVPTAO     ................C...........A.....AG.G...........................C..  2234
GBVPTRL     .....................T...................................C...........  2231
GBVPTTH     .  ........AA.AA........................................CA.......C.....  2240

MAJORITY   GCCCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCGGACCTCATGAAGCTGGCCATGGTGAAGCTC

GBVPTAO     .    ..............................................T............  2304
GBVPTRL        ......'............G.....................................CG.  .T  2301
GBVPTTH     .   ........................................................C...........    .  2310

MAJORITY   TTCCCCCGGCTXCAGGAAATGGGGGCCAGGATGCTCCTXCAGGTCCACGACGAGCTGGTCCTCGAGGCCC

GBVPTAO          A ...GG........................T..............................  .  2374
GBVPTRL             ...T......C.......G........TT.G.....G.........................  2371
GBVPTTH    .      ...C..C.G...G.........C.C........C.................CC....G.......  2380

MAJORITY   CCAAAGACCGGGCGGAGGXGGTGCCCCGCTTTGGCCAAGGAGGTCATGGAGGGGGTCTATCCCCTGGCCGT

GBVPTAO    .A  .....A..........CC......CGGC...........................G..............  2444
GBVPTRL        .G..C..........AG...A.....................................GG......CAG..  2441
GBVPTTH      .C   .C.......C....A.......G.................C.......AA..C..........C.....  2450
```

FIGURE 1H

```
MAJORITY   GCCCCTGGAGGTGGAGGTGGGGATGGGGGAGGACTGGCTCTCCGCCAAGGAGTAG

INJPTAQ    ........................A............................GA   2499
INJPTFL    .....................CC................................   2496
INJPTTH    .......................................T..........GT...   2505
```

FIGURE 2A

```
[illegible]  MXAML PLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDG DAVXVVFDAK

[illegible]  RG..........................H..............................I.....  69
[illegible]  ...........................................................V.V......  68
[illegible]  E..........................................................YK..F......  70

[illegible]  APSFRHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGLXRLEVPGYEADDVLATLAKKAEKEGYEVRIL

[illegible]  ........GG........................A..............S...............  139
[illegible]  ..................................V......F...........R...........  138
[illegible]  .................................FT...............................  140

[illegible]  TADRQLYQLLSDRIAVLHPEGYLITPAWLWEKYGLRPEQWVDYRALXGDPSDNLPGVKGIGEKTAXKLLX

[illegible]  .K.........H......................D..A.....T..E...............R.  E  209
[illegible]  ......E...I..............Y................A......I...........QR.IR  209
[illegible]  .V...V.....H....E................F..V.................L     K  210

[illegible]  EWGSLENLLKNLDRVKP XXREKIXAHMEDLXLSXXLSXVRTDLPLEVDFAXRREPDREGLRAFLERLEF

[illegible]  A.....L..AI....L...D..K..WD.AK............K.......R..........  278
[illegible]  ..FQH..Q....SL...LQ.G..A.A..RK..Q.H.........GR..T.HL............  277
[illegible]  .......EHV....K..L...R..LE..R..........L.QG................  280

[illegible]  GSLLHEFGLLEXPKALEEAPWPPPEGAFVGFVLSRPEPMWAELLALAAARXGRVHRAXDPLXGLRDLKEV

[illegible]  ...........S.......................K.....D........G......PE.YKA......A  348
[illegible]  ...........G...A.............L..SF..............G.WE..L...D...R......G.  347
[illegible]  ..........A.AP............................K....G.D......A...A..K.....  350
```

FIGURE 2B

```
MAJORITY  RGLLAKDLAVLALREGLDLXPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEDAGERALLSERLFXNLXX

TAQ PRO   .......S.......G.P...............................E.....A......A..WG  418
TFL PRO   I..............F.E......................................A....QT.KE  417
TTH PRO   ..........S......V...................................AH........HR..LK  420

MAJORITY  RLEGEERLLWLYXEVEKPLSRVLAHMEATGVRLDVAYLOALSLEVAEEIRRLEEEVFRLAGHPFNLNSRD

TAQ PRO   ..........R...R...A.................R..........A...A................  488
TFL PRO   K.........E...........R....................EA.V.Q...................  487
TTH PRO   ....K.....H........................L...................................  490

MAJORITY  QLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKNTYIDPLPXLVHPRTG

TAQ PRO   ..............................................S.......D.I.....  558
TFL PRO   ......................................DR...................A....K..  557
TTH PRO   ....R...L...Q.........................H..........V....S.......  560

MAJORITY  RLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFVAEEGWXLVALDYSQIELRVLAHLSGDENL

TAQ PRO   ...................................I.....L.....................  628
TFL PRO   .........................................V..V....................  627
TTH PRO   ......................................A..A.......................  630

MAJORITY  IRVFQEGRDIHTQTASWMFGVPPEAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAVAFIERYFQ

TAQ PRO   ......E.........R......................................Q........  698
TFL PRO   ..................S..G..........................G..S...............  697
TTH PRO   ..K........................V..................................  700
```

FIGURE 2C

```
MAJORITY  SFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLNARVKSVREAAERMAFNMPVQGTAADLMKLAMVKL

TAQ PRO   ...................................E..................................  768
TFL PRO   .Y........G.........................................................R.  767
TTH PRO   .................K....................................................  770

MAJORITY  FPRLXEMGARMLLQVHDELVLEAPKXRAEXVAALAKEVMEGVYPLAVPLEVEVGXGEDWLSAKEX

TAQ PRO   ....E....................E...A..R.....................I.........        833
TFL PRO   ....Q.L..................D...R............W..Q........L.........        831
TTH PRO   ....R..............L....QA...E.......A..KA............M........G        835
```

FIGURE 3

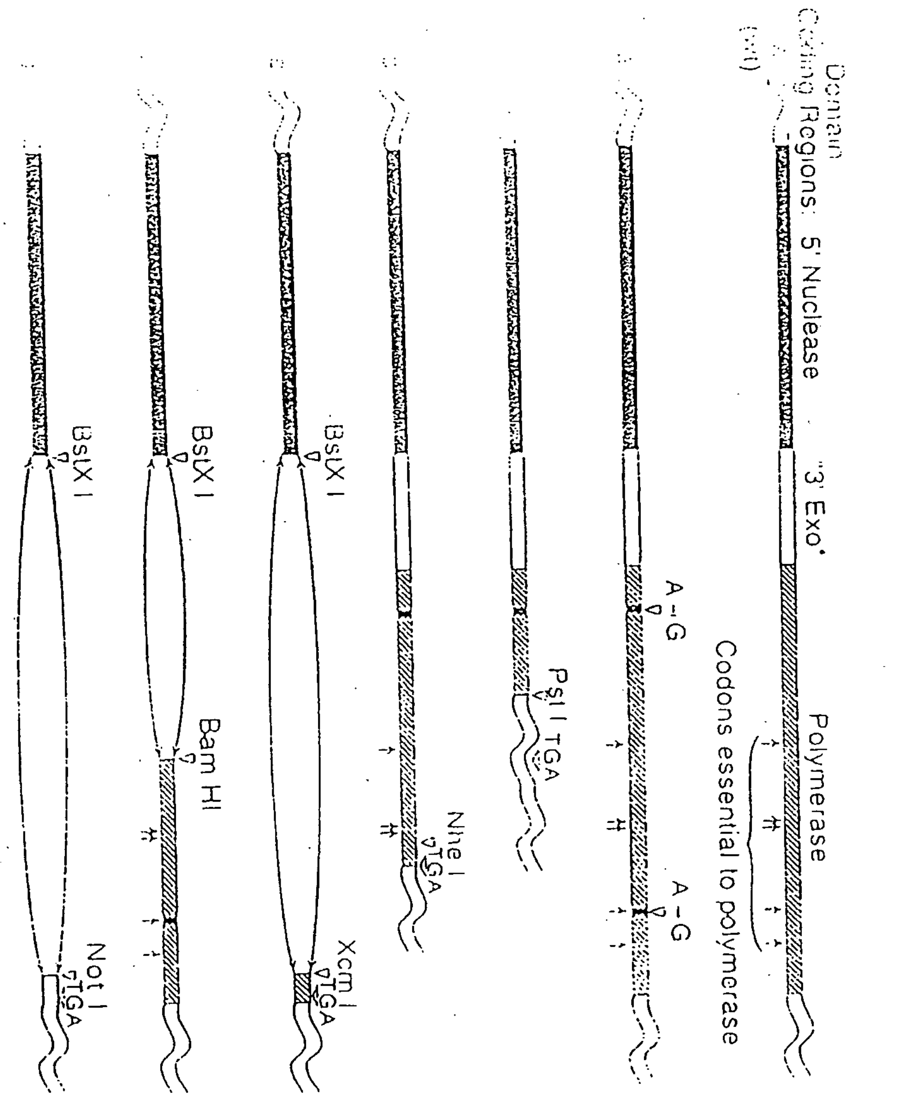
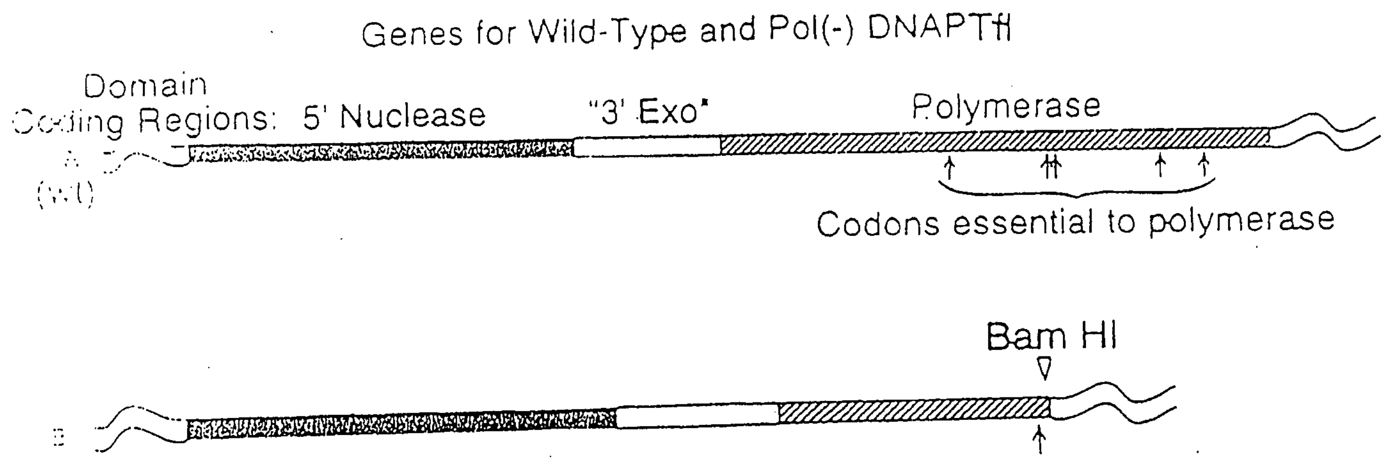

Domain
Coding Regions: 5' Nuclease
"3' Exo"
Polymerase
Codons essential to polymerase
A→G
A→G
PstI TGA
Nhe I
TGA
BstX I
Xcm I
TGA
BstX I
Bam HI
BstX I
Not I
TGA

FIGURE 4


Genes for Wild-Type and Pol(-) DNAPTfl
Domain
Coding Regions:
5' Nuclease
"3' Exo"
Polymerase
A
(wt)
Codons essential to polymerase
B
Bam HI

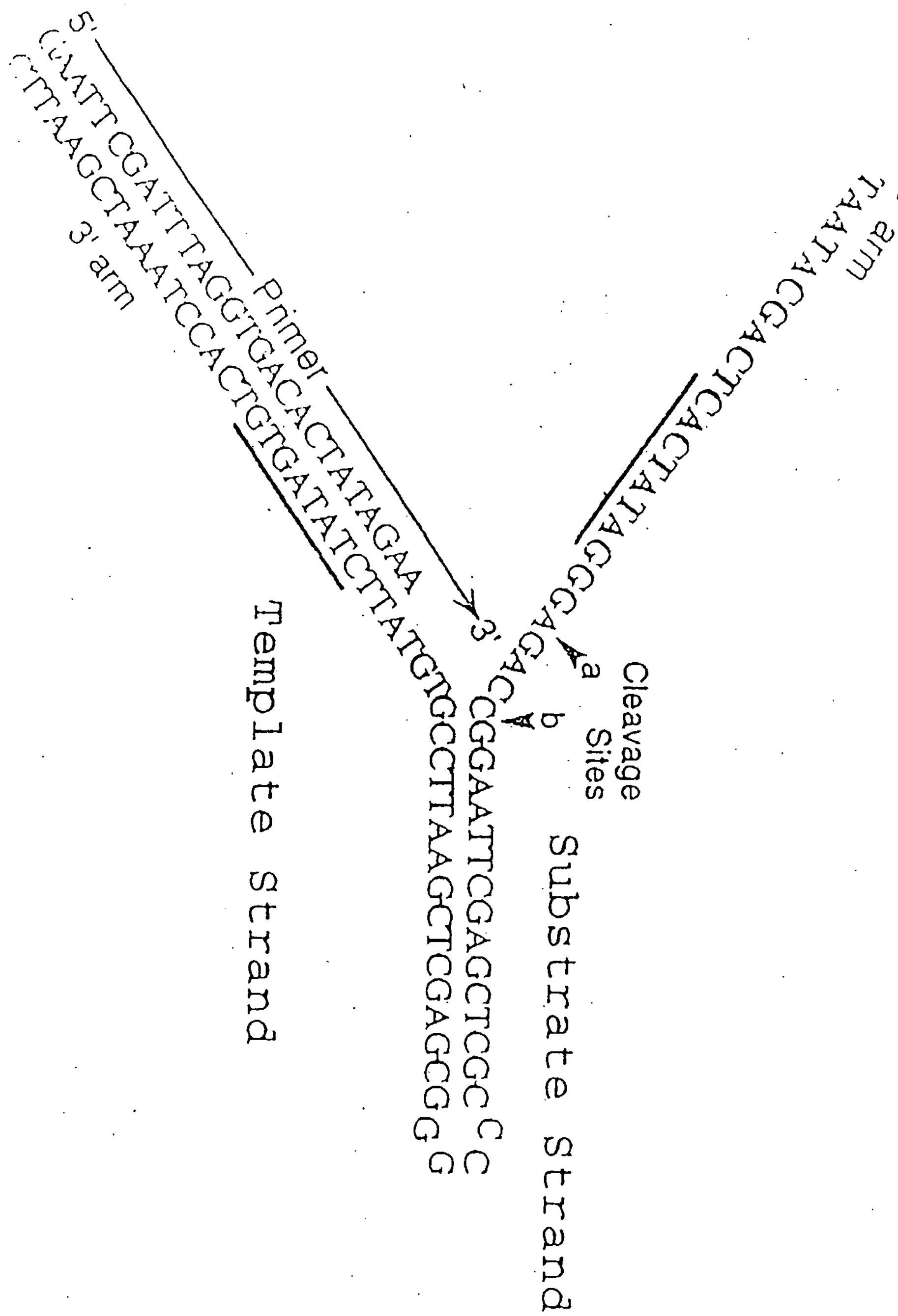

5'
Primer
3' arm
5' arm
3'
Cleavage Sites
a
b
Substrate Strand
Template Strand

FIGURE 5

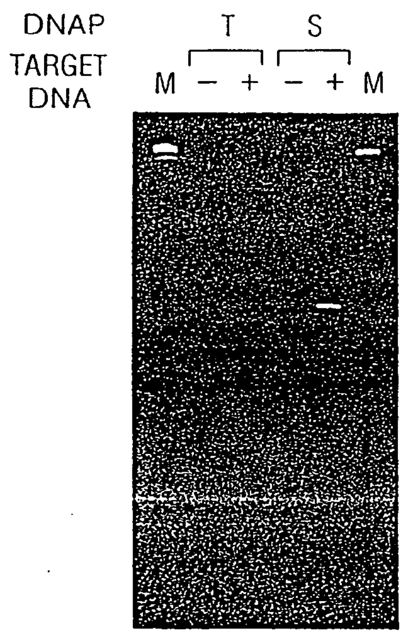


FIGURE 6

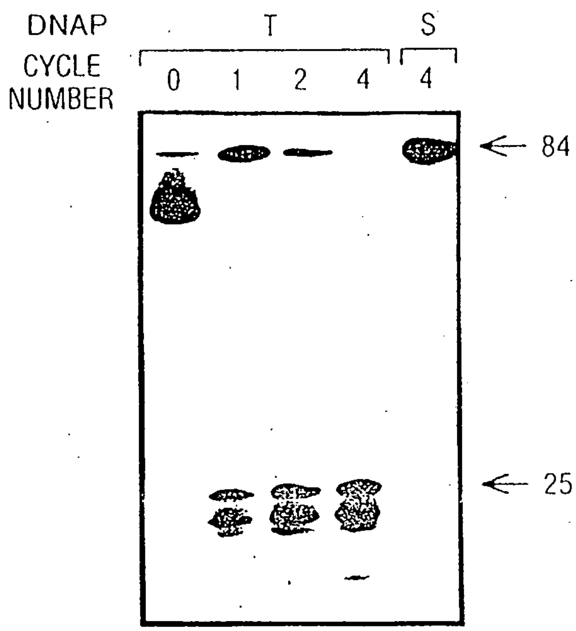


FIGURE 7

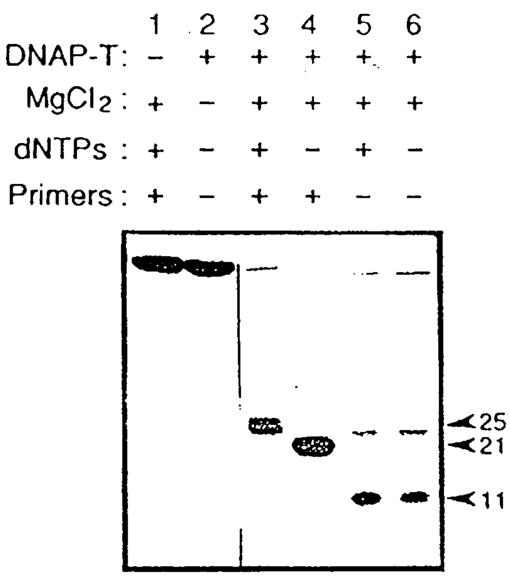


FIGURE 8A

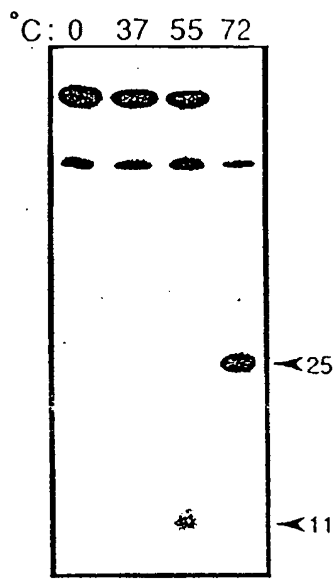


FIGURE 8B

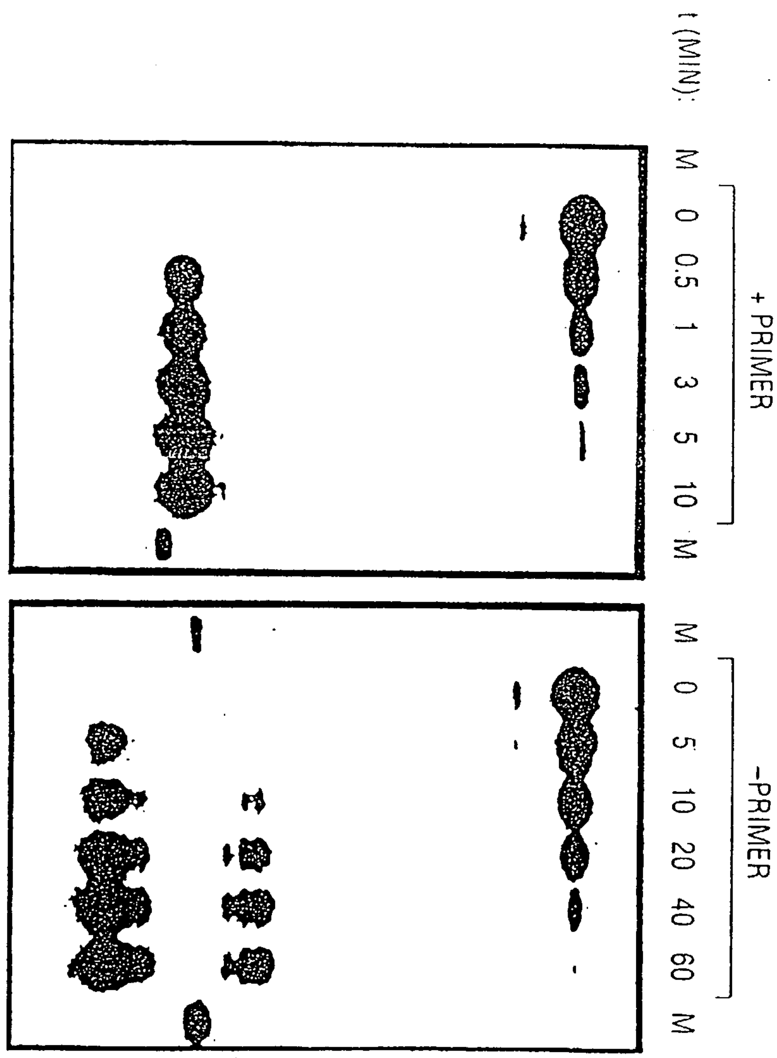


FIGURE 9A

FIGURE 9B

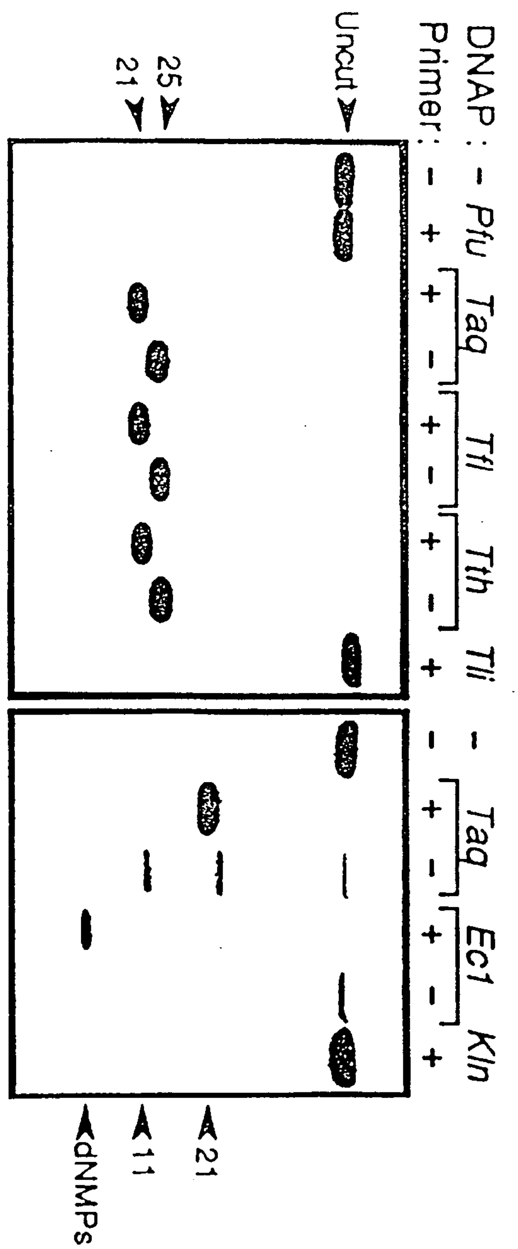


FIGURE 10A

FIGURE 10B

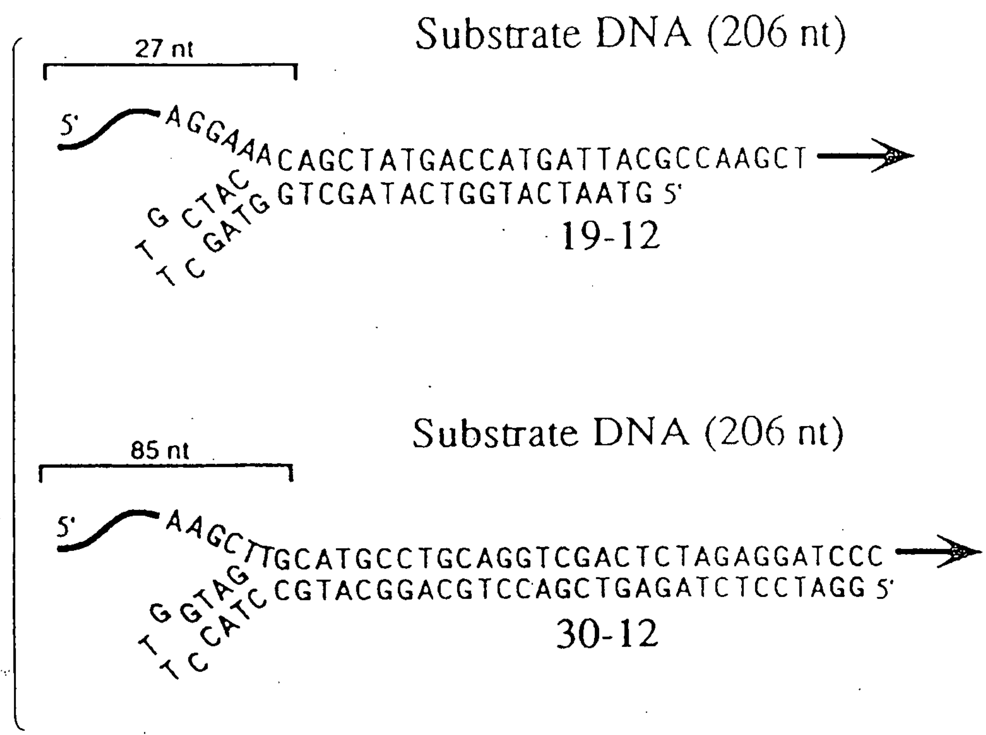


FIGURE 11A

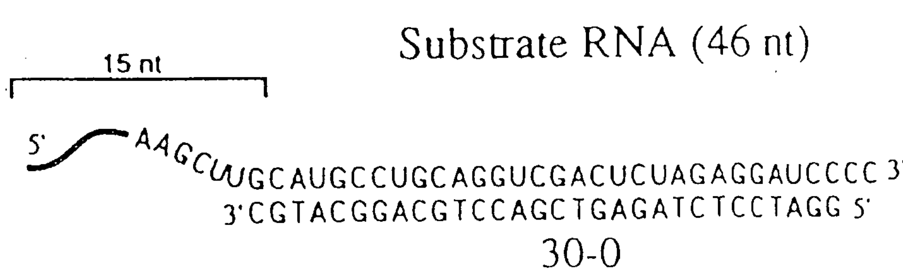


FIGURE 12A

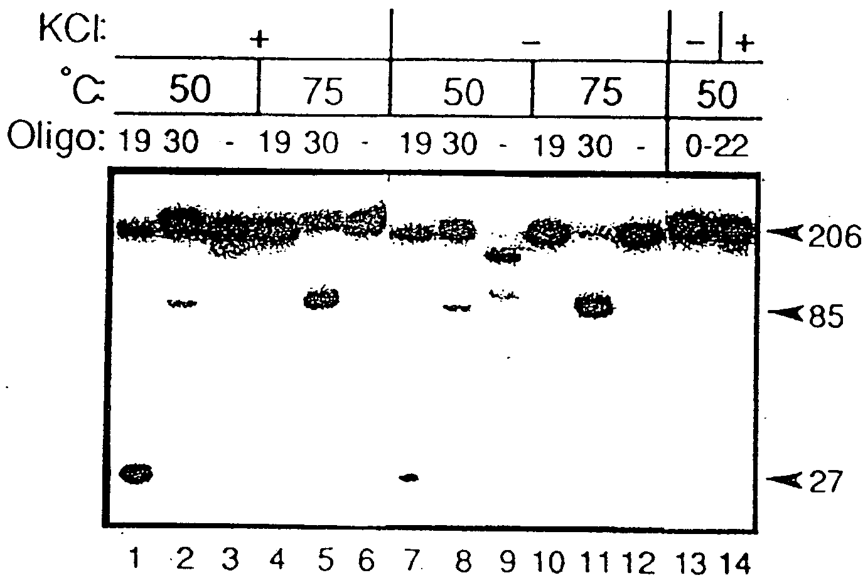


FIGURE 11B

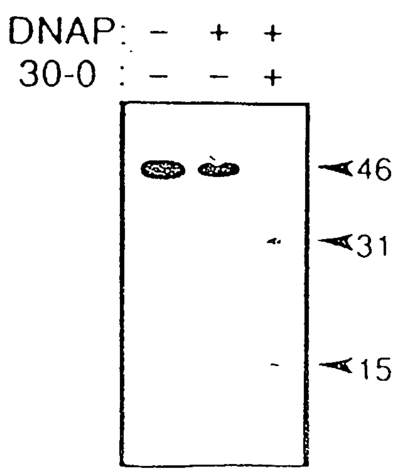
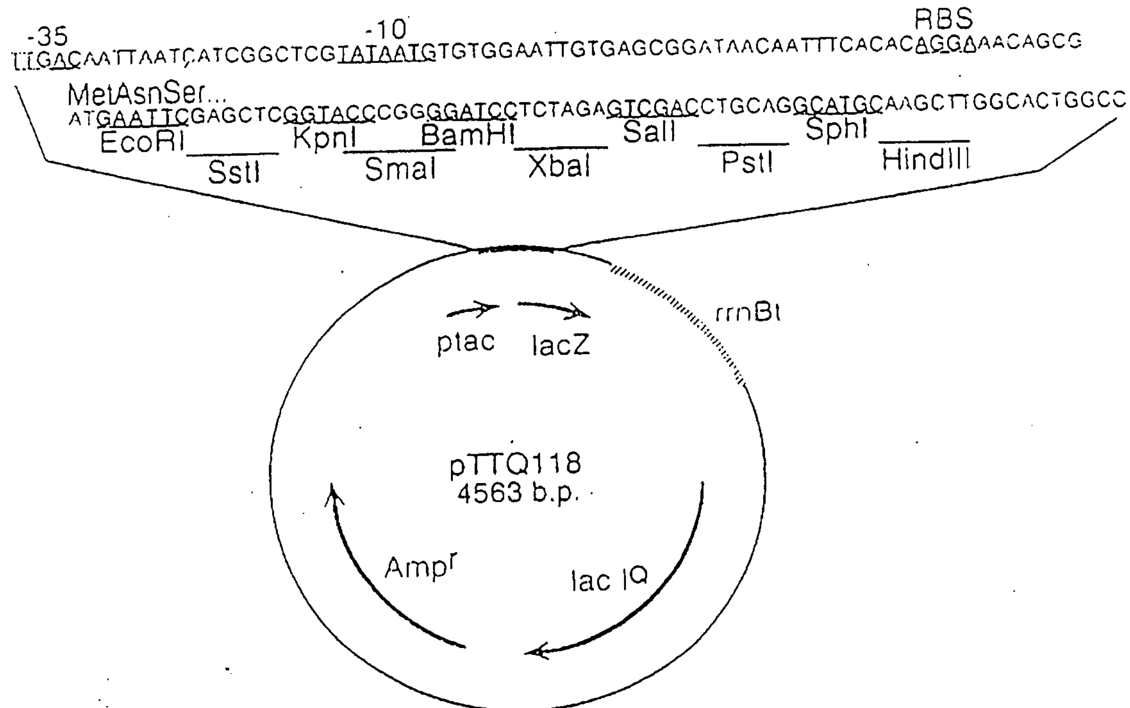


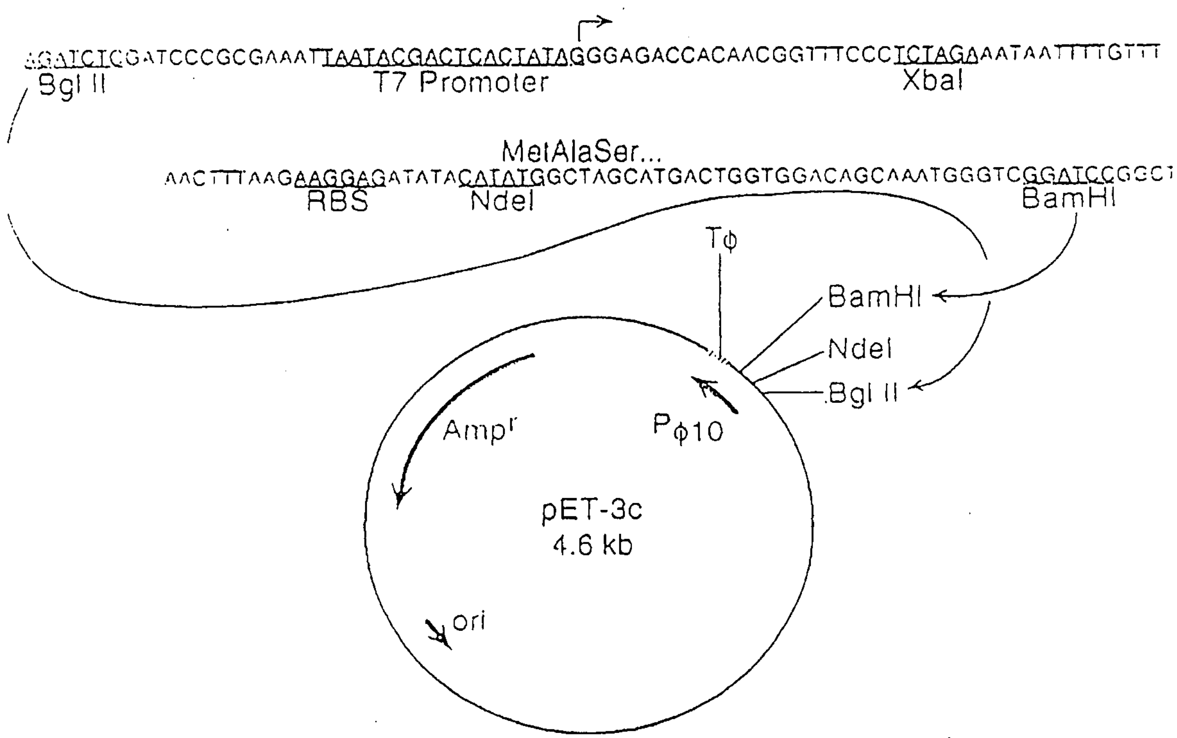
FIGURE 12B

FIGURE 13



RBS: Ribosome binding site
ptac: Synthetic tac promoter
lac IQ: Lac repressor gene
lacZ: Beta-galactosidase alpha fragment
rrnBt: E. coli rrnB transcription terminator

FIGURE 14



Pφ10: Bacteriophage T7 φ10 promoter
Tφ: T7 φ Terminator
RBS: Ribosome binding site

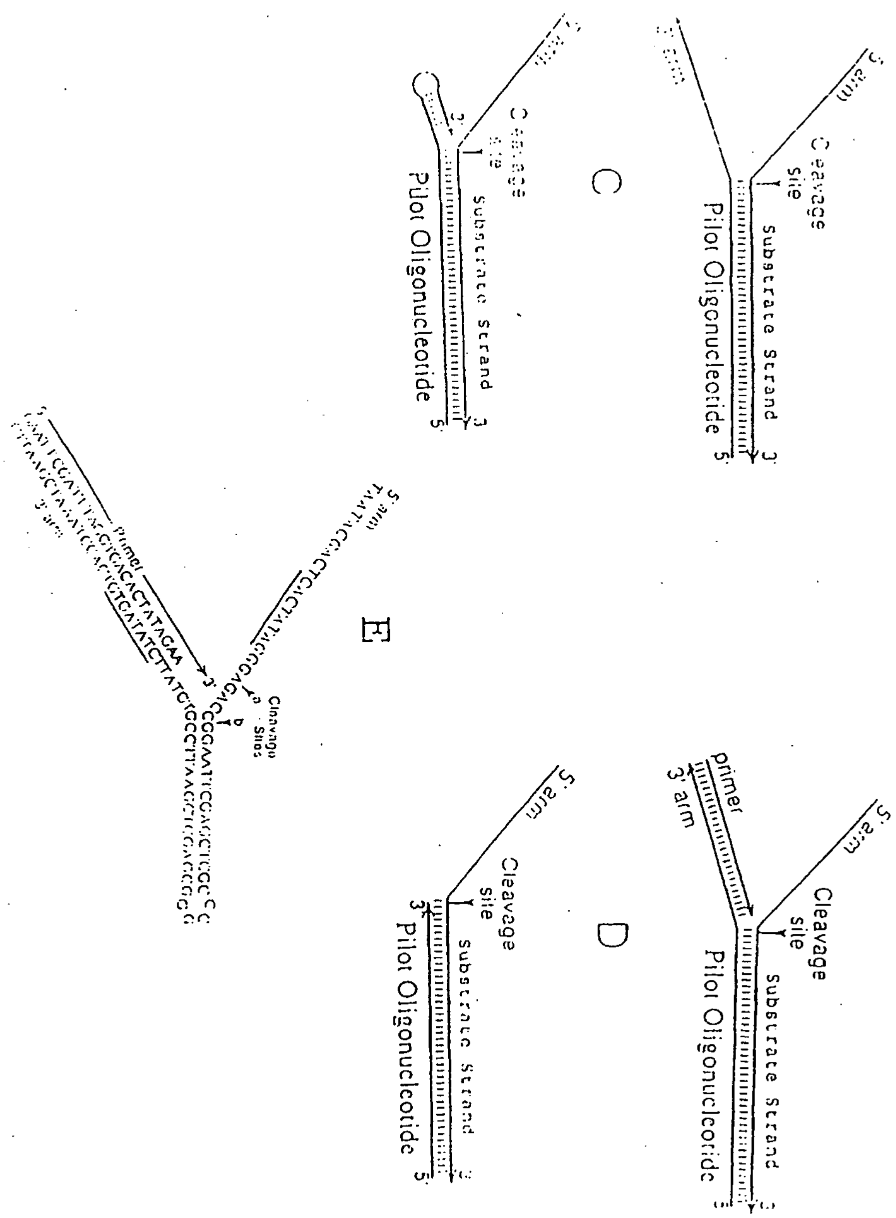

A
B
C
D
E
Cleavage site
Substrate Strand
Pilot Oligonucleotide
5' arm
3' arm
primer
Primer
Cleavage Sites

FIGURE 15

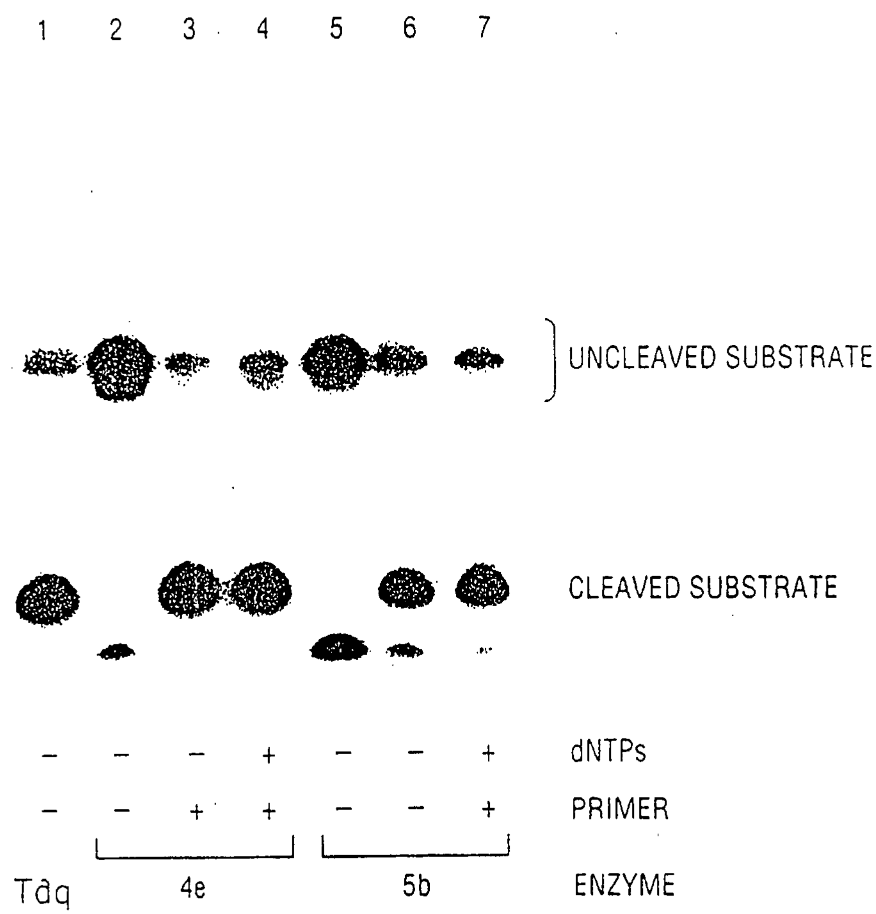

1 2 3 4 5 6 7
UNCLEAVED SUBSTRATE
CLEAVED SUBSTRATE
− − − + − − + dNTPs
− − + + − − + PRIMER
Taq 4e 5b ENZYME

FIGURE 16

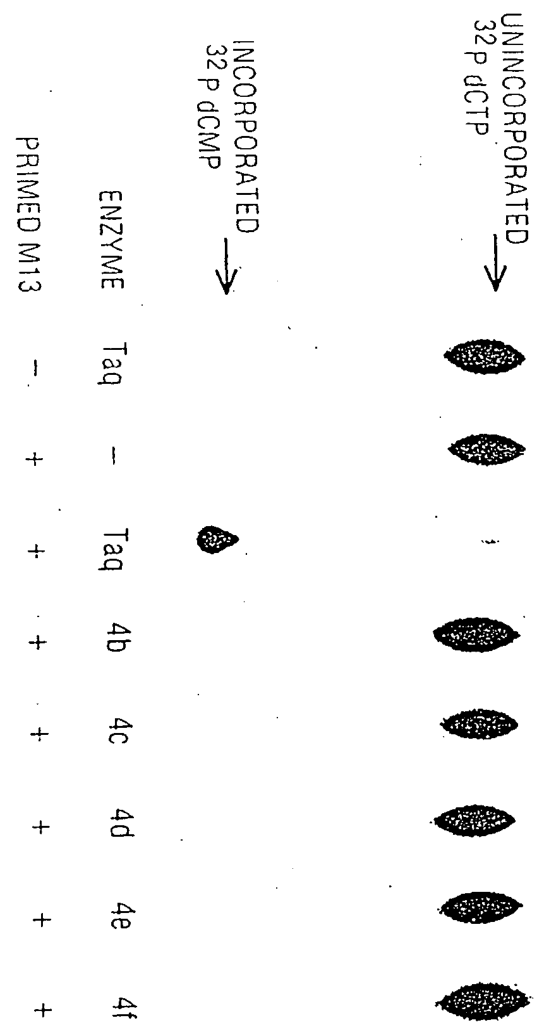

UNINCORPORATED
32 P dCTP
INCORPORATED
32 P dCMP
ENZYME
Taq
–
Taq
4b
4c
4d
4e
4f
PRIMED M13
–
+
+
+
+
+
+
+

FIGURE 17

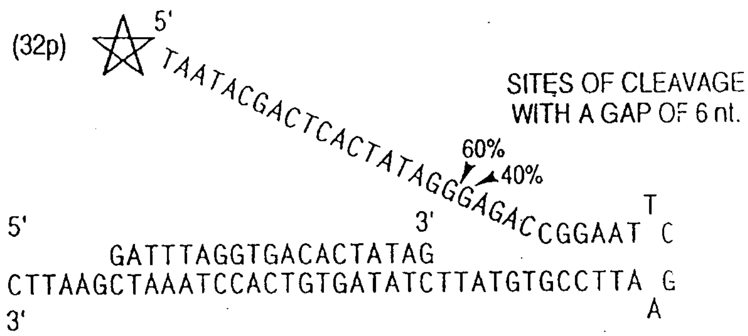

(32p)
5'
TAATACGACTCACTATAGGGAGACCGGAAT
SITES OF CLEAVAGE WITH A GAP OF 6 nt.
60%
40%
5'
3'
GATTTAGGTGACACTATAG
CTTAAGCTAAATCCACTGTGATATCTTATGTGCCTTA
3'
T
C
G
A

FIGURE 18A

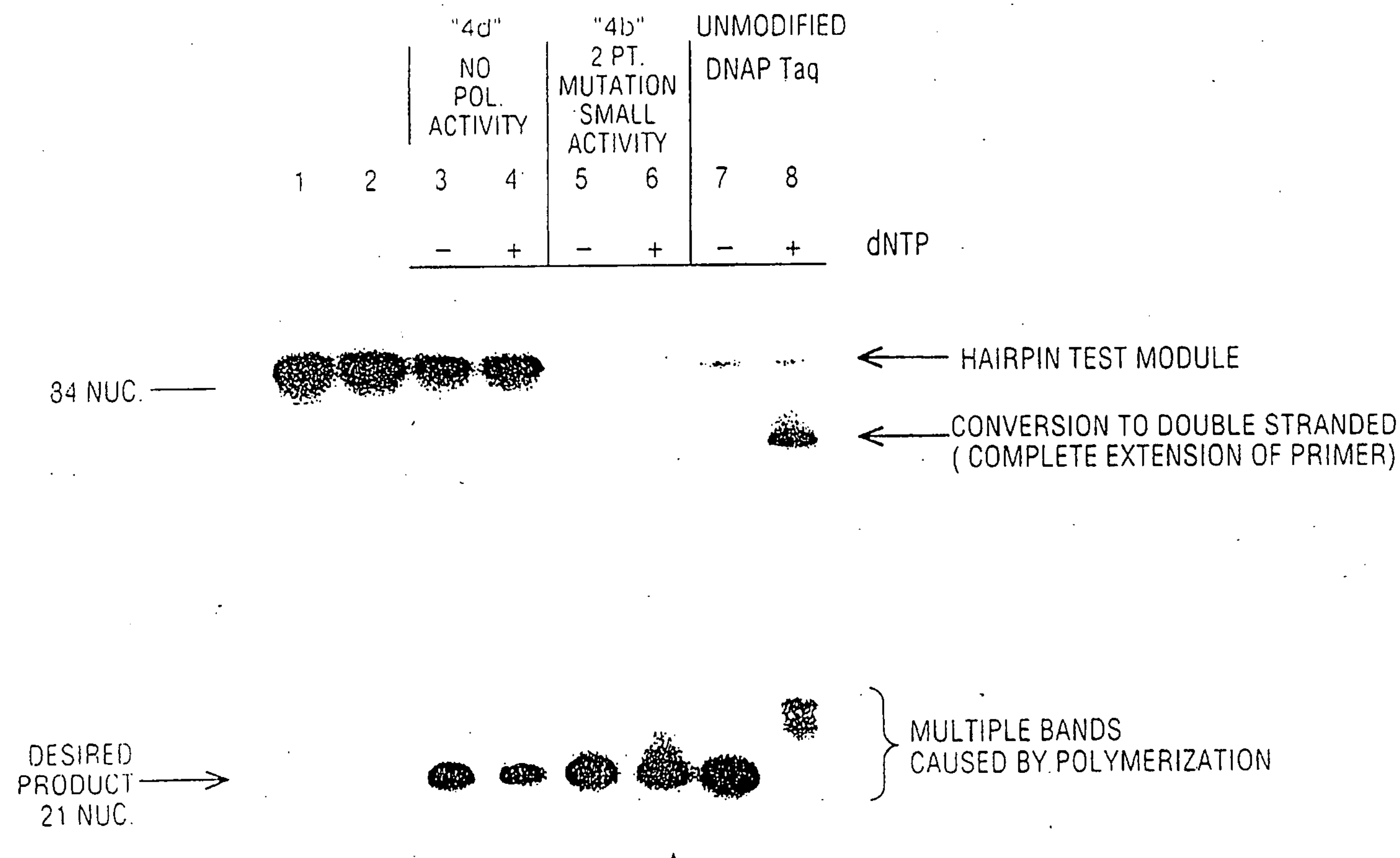

"4d"
NO
POL.
ACTIVITY
"4b"
2 PT.
MUTATION
SMALL
ACTIVITY
UNMODIFIED
DNAP Taq
1
2
3
4
5
6
7
8
–
+
–
+
–
+
dNTP
34 NUC.
HAIRPIN TEST MODULE
CONVERSION TO DOUBLE STRANDED
(COMPLETE EXTENSION OF PRIMER)
DESIRED
PRODUCT
21 NUC.
MULTIPLE BANDS
CAUSED BY POLYMERIZATION
SOME ABERRANT CLEAVAGE WITH "4b"
BECAUSE OF RESIDUAL POLYMERASE ACTIVITY.

FIGURE 18B

FIGURE 19

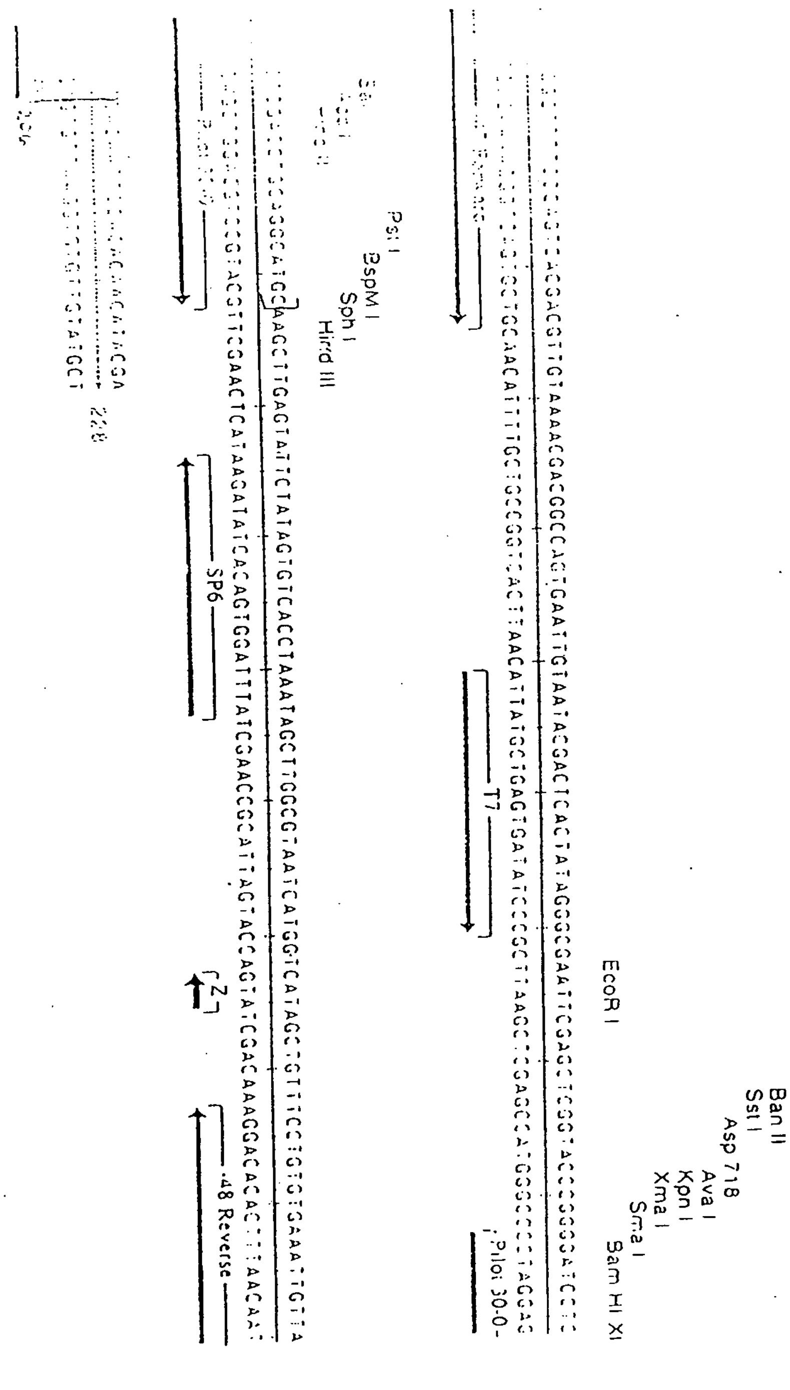
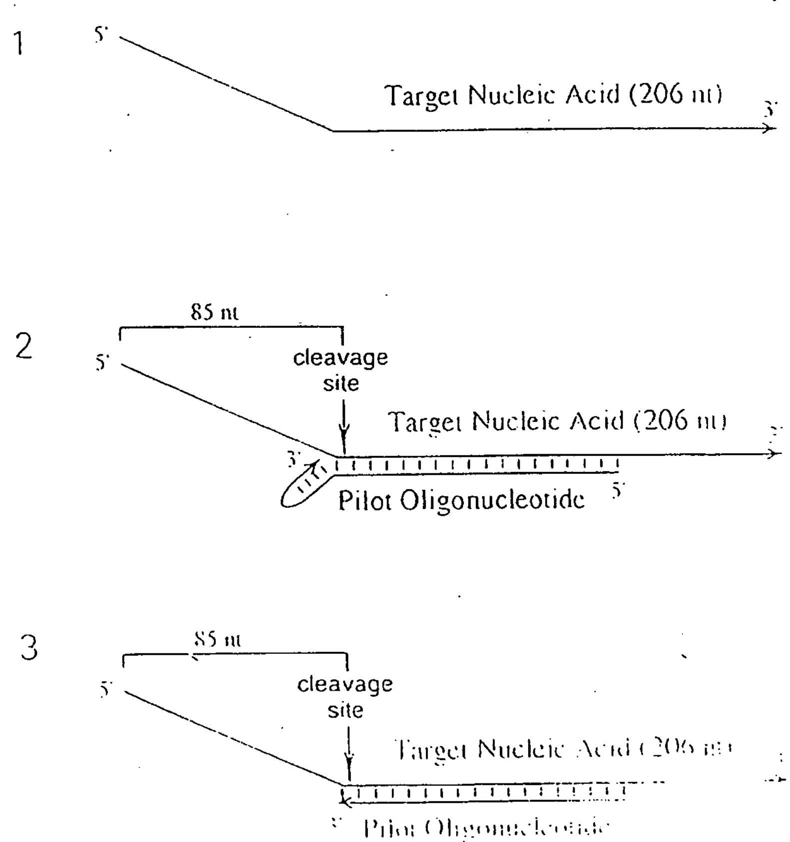

Ban II
Sst I
Asp 718
Ava I
Kpn I
Xma I
Sma I
Bam HI XI
EcoR I
Pilot 30-0
T7
SP6
-48 Reverse
Pst I
BspM I
Sph I
Hind III
228

FIGURE 20A


1
5'
Target Nucleic Acid (206 nt)
3'
2
85 nt
5'
cleavage
site
Target Nucleic Acid (206 nt)
3'
Pilot Oligonucleotide
5'
3
85 nt
5'
cleavage
site
Target Nucleic Acid (206 nt)
Pilot Oligonucleotide

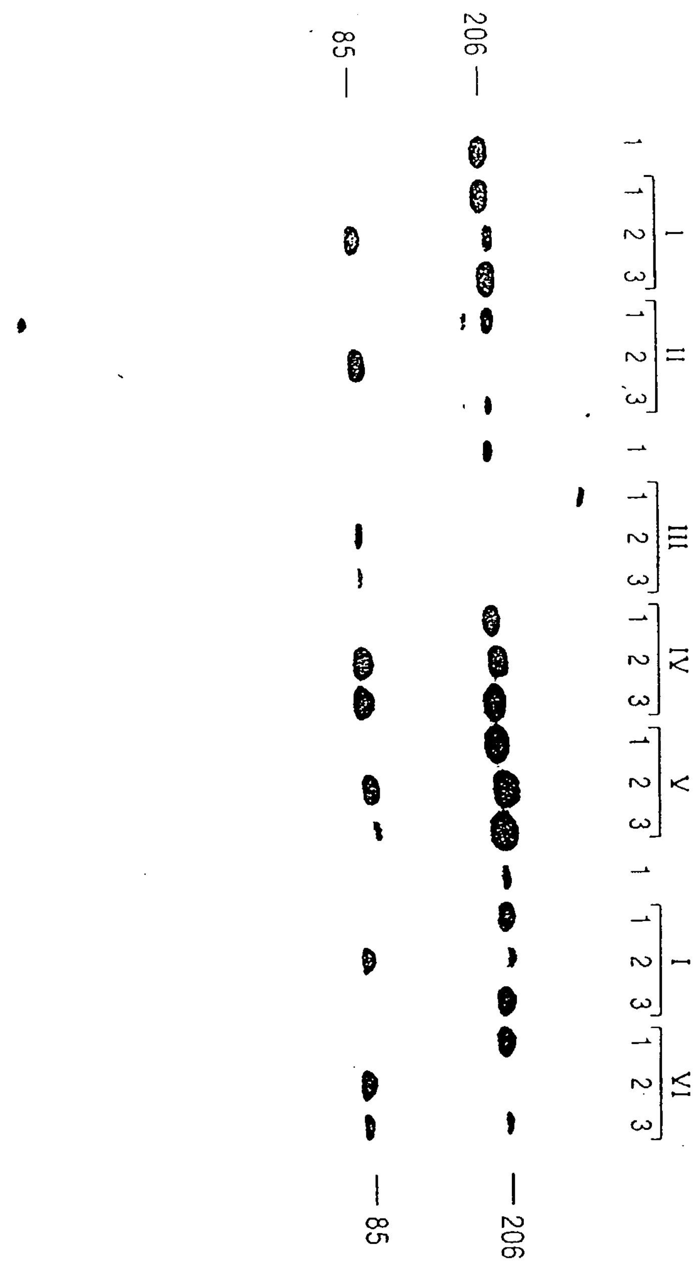


FIGURE 20B

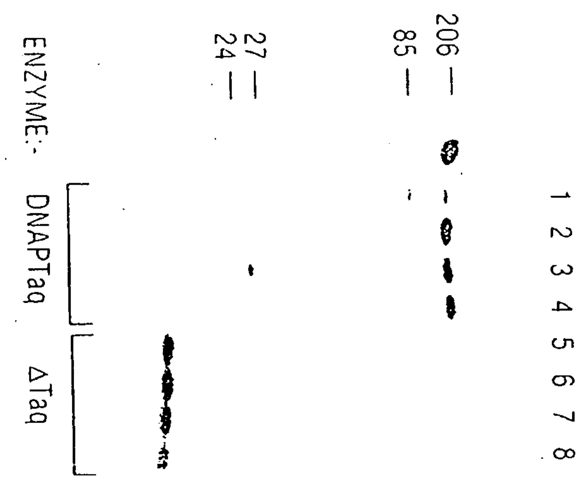


FIGURE 21A

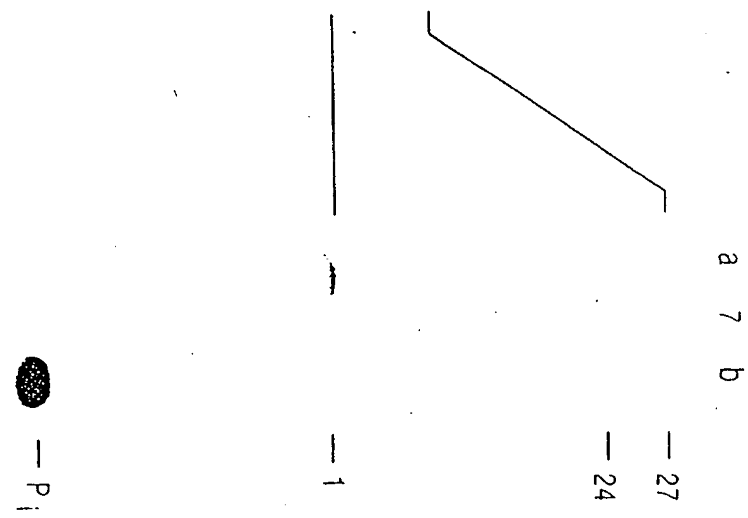
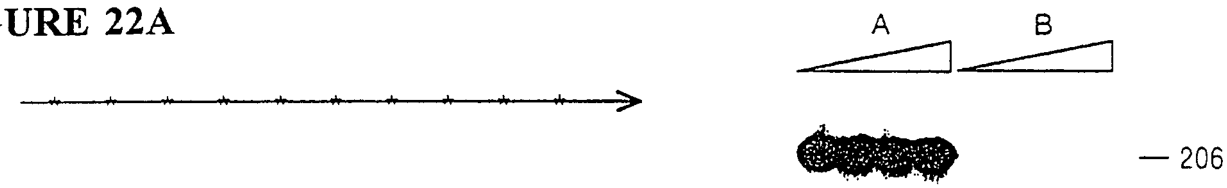


FIGURE 21B

FIGURE 22A



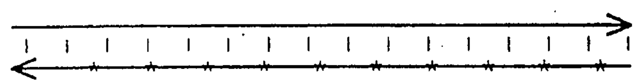
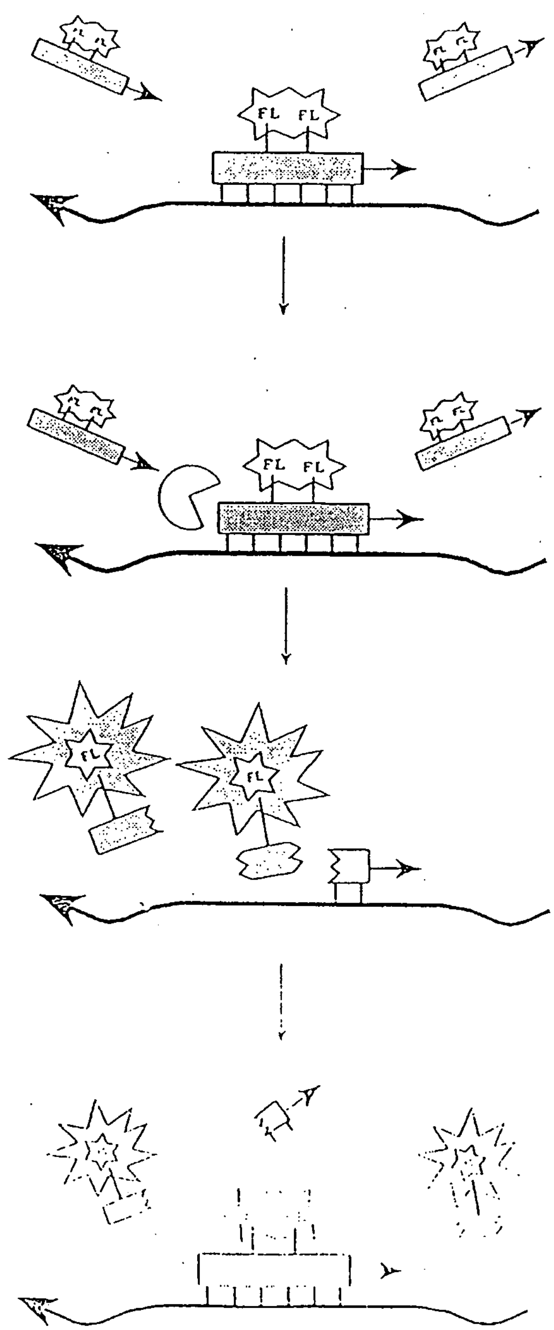


FIGURE 22B

* = 32p

FIGURE 23



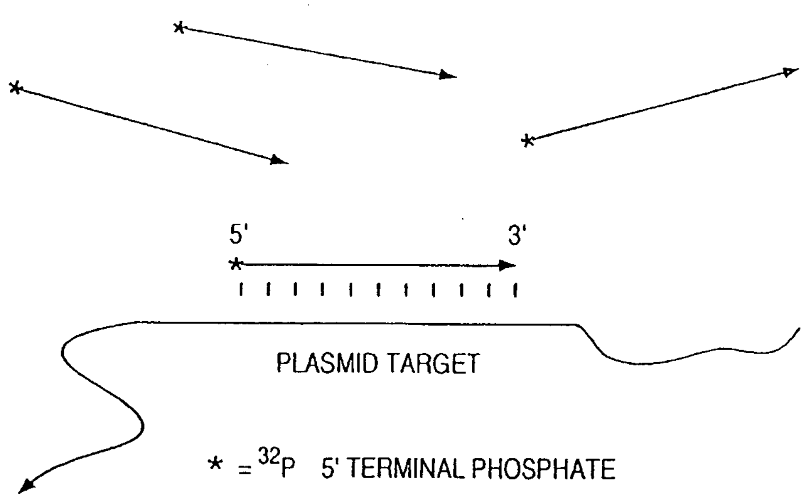


FIGURE 24A

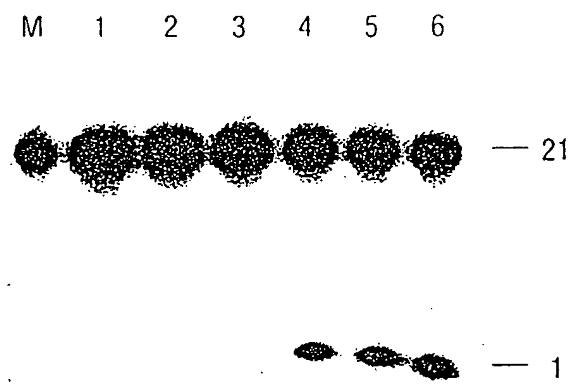


FIGURE 24B

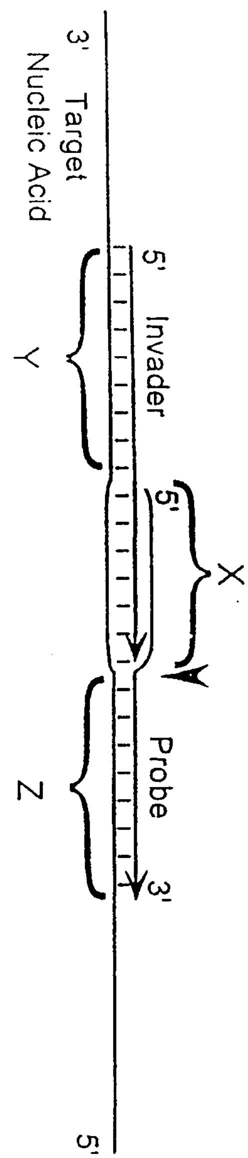
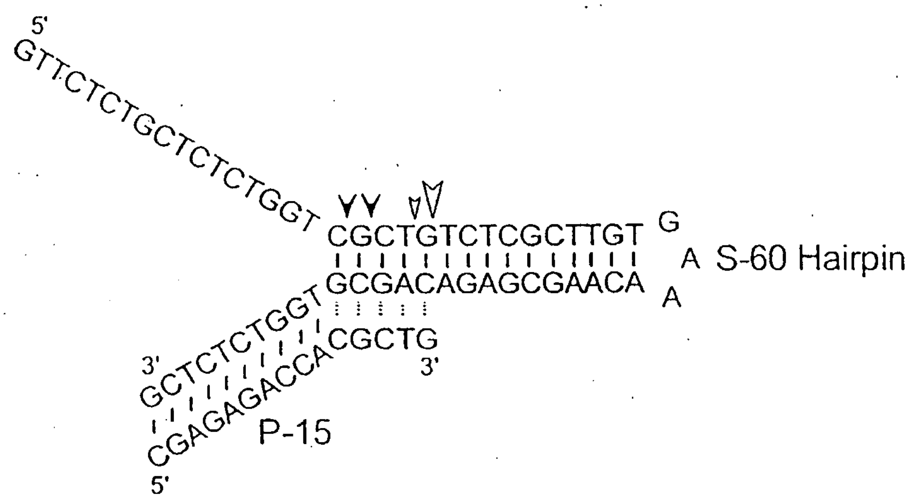


FIGURE 25

FIGURE 26


5'
GTTCTCTGCTCTCTGGT
CGCTGTCTCGCTTGT
G
A
A
S-60 Hairpin
GCGACAGAGCGAACA
GCTCTCTGGT
CGCTG
3'
3'
CGAGAGACCA
P-15
5'

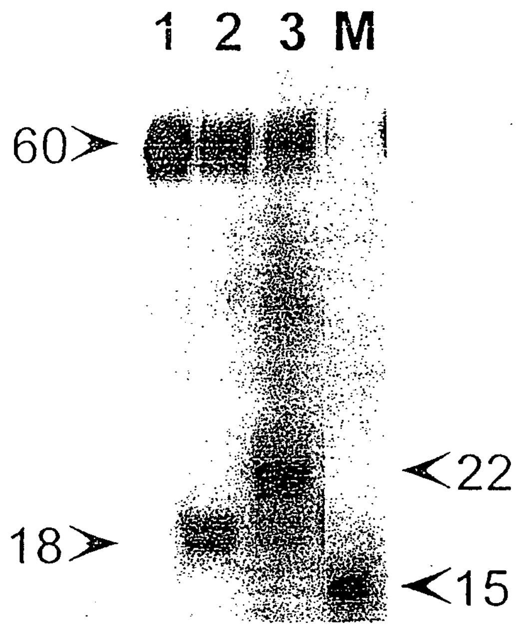
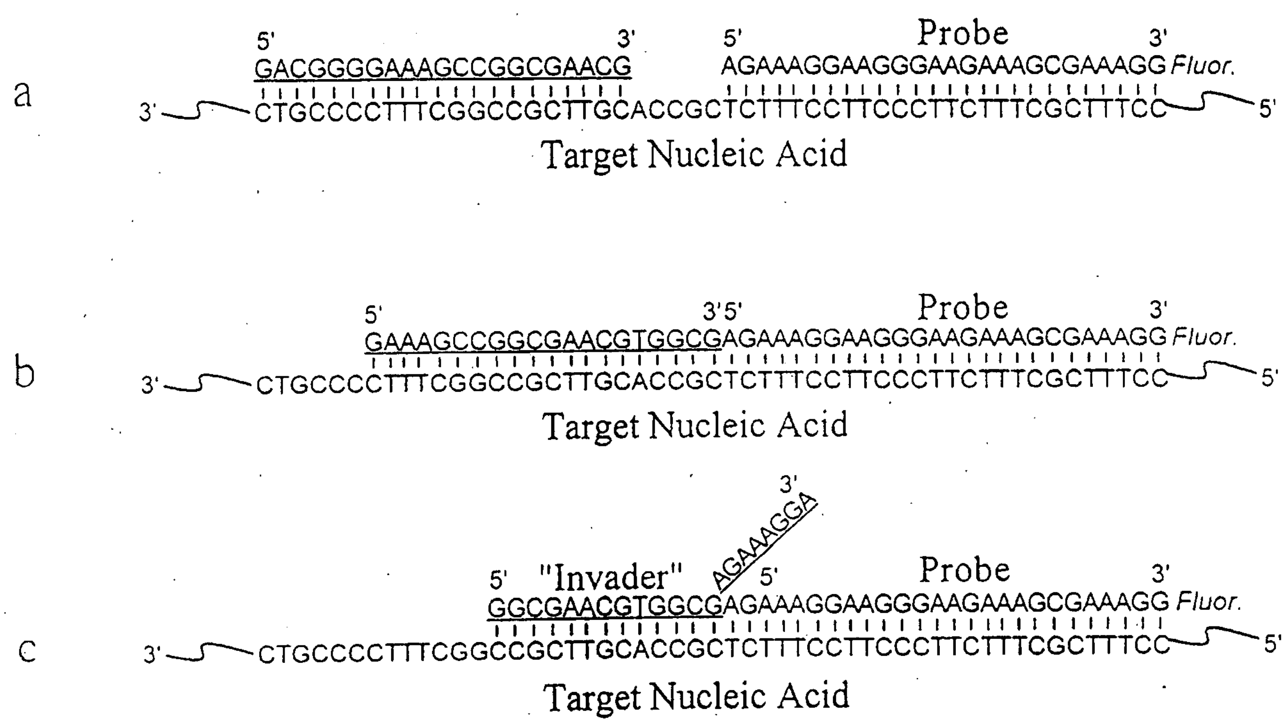


FIGURE 27

FIGURE 28


a
5'
GACGGGGAAAGCCGGCGAACG
3'
5'
Probe
3'
AGAAAGGAAGGGAAGAAAGCGAAAGG Fluor.
3'
CTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCC
5'
Target Nucleic Acid
b
5'
GAAAGCCGGCGAACGTGGCG
3'5'
Probe
3'
AGAAAGGAAGGGAAGAAAGCGAAAGG Fluor.
3'
CTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCC
5'
Target Nucleic Acid
c
3'
AGAAAGGA
5'
"Invader"
GGCGAACGTGGCG
5'
Probe
3'
AGAAAGGAAGGGAAGAAAGCGAAAGG Fluor.
3'
CTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCC
5'
Target Nucleic Acid

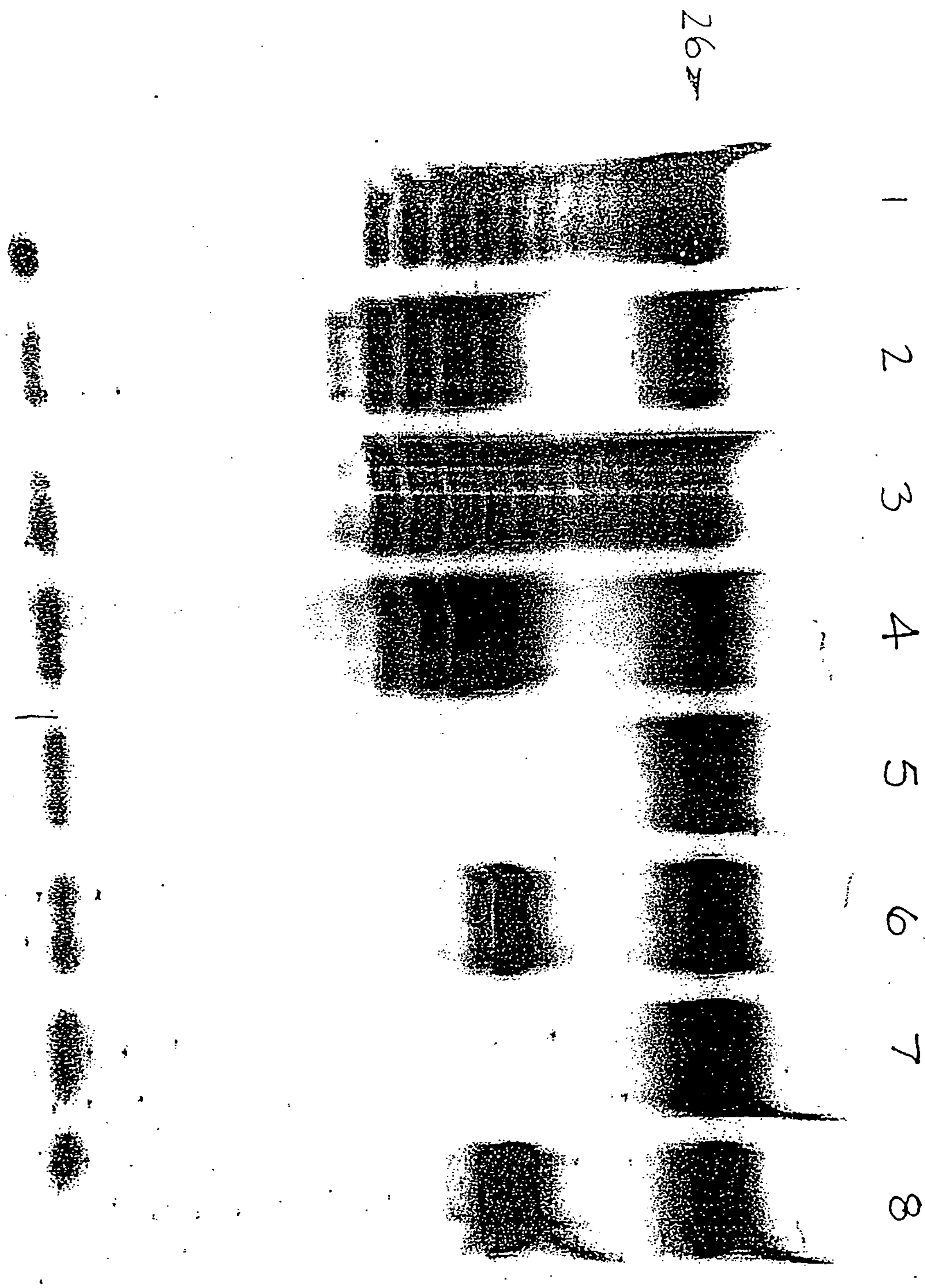


FIGURE 29

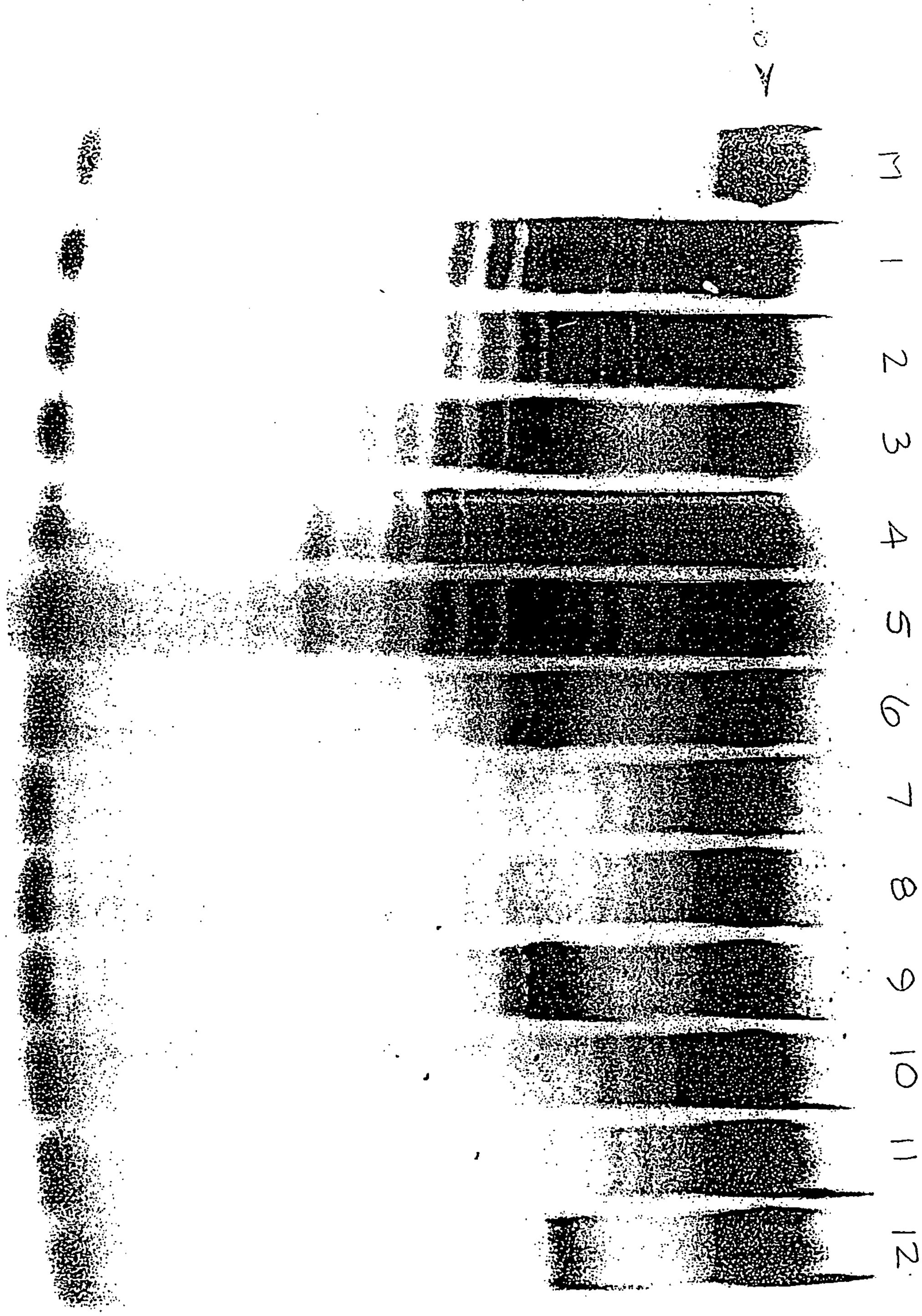


FIGURE 30

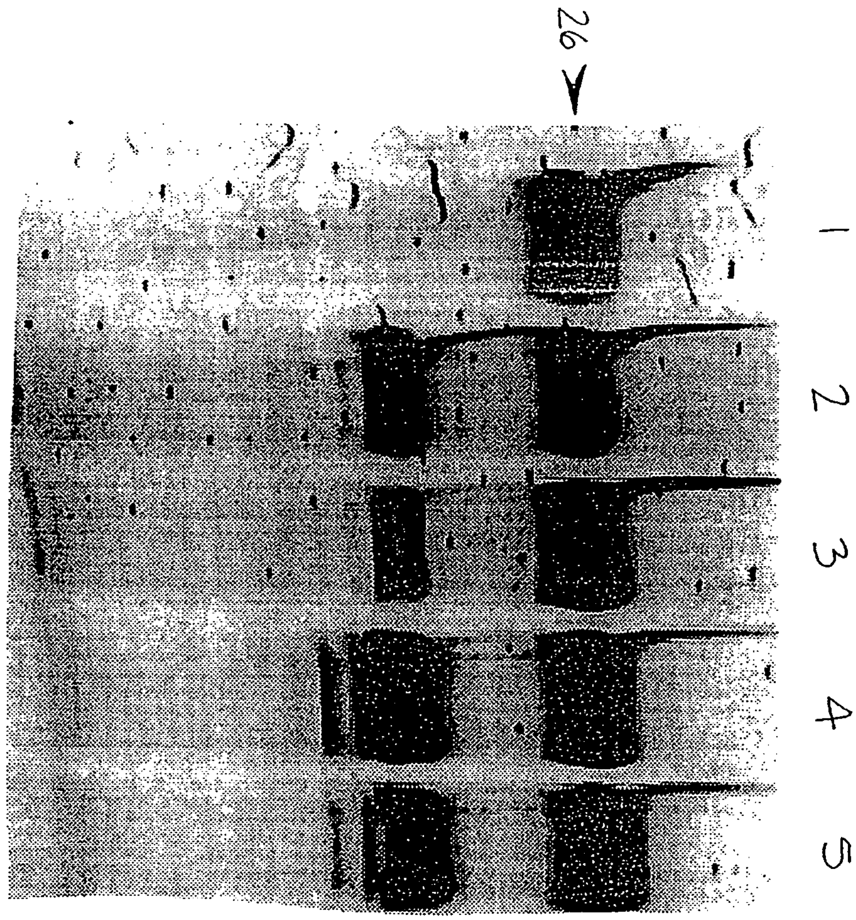
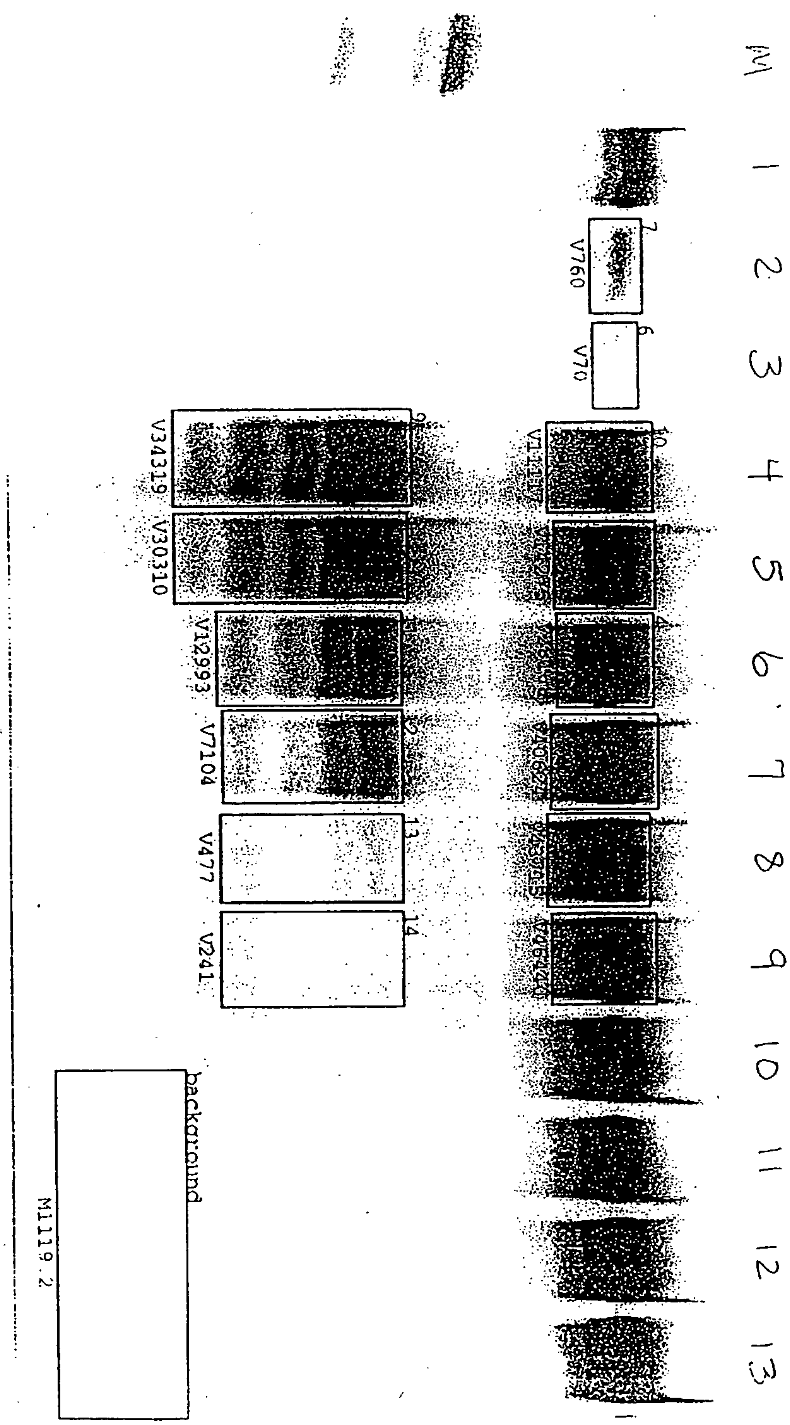


FIGURE 31

FIGURE 32


M
1
2
3
4
5
6
7
8
9
10
11
12
13
V760
V70
V34319
V30310
V12993
V7104
V477
V241
background
M1119.2

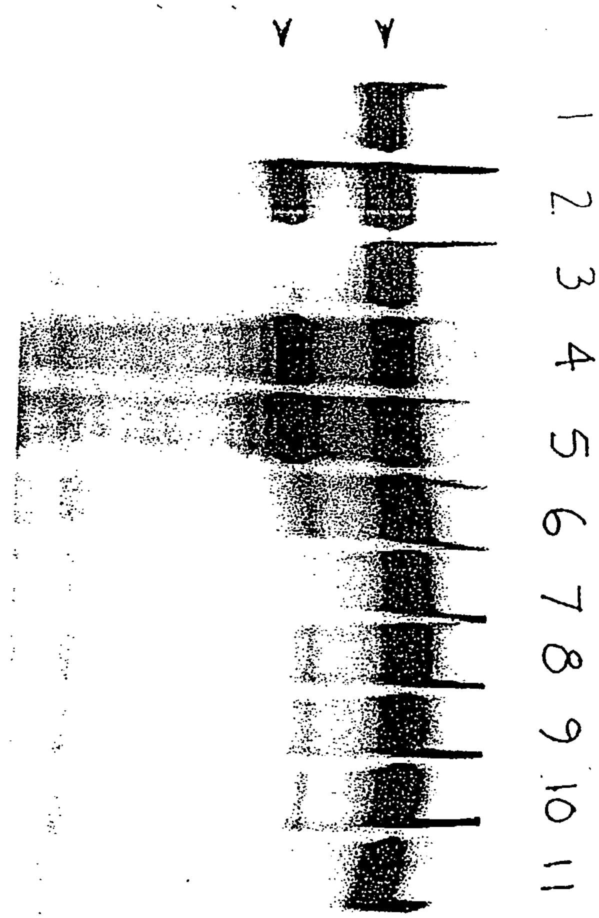


FIGURE 33

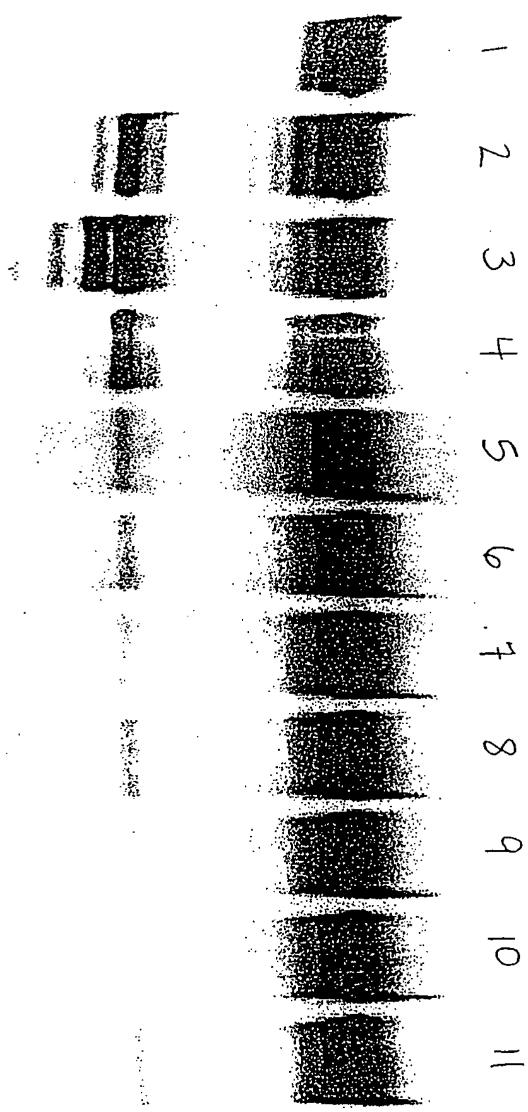
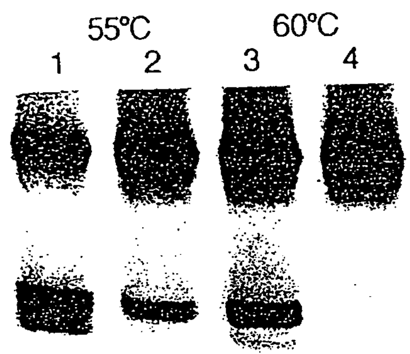


FIGURE 34

FIGURE 35



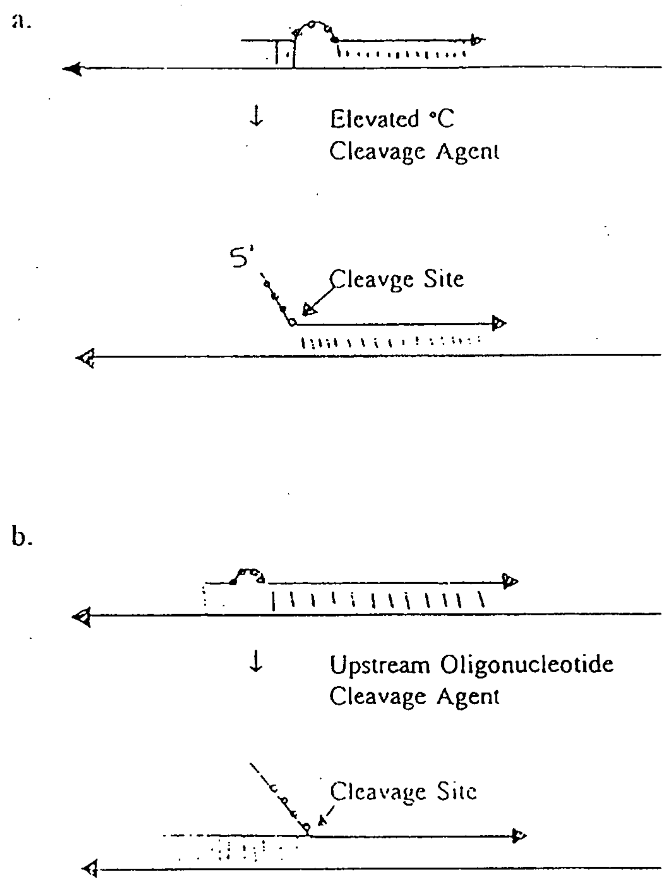

a.
Elevated °C
Cleavage Agent
5'
Cleavge Site
b.
Upstream Oligonucleotide
Cleavage Agent
Cleavage Site

FIGURE 36

```
                                          Cleavage Site
Invasive Cleavage Directing Oligo               ↓
                       5'-cy-3agaaaggA agggaagaaagcgaaaggt-3'   Probe Oligomer
         5'-gccggcgaacgtggcgagaaagga
3'-ctgcccctttcggccgcttgcaccgctctttccttcccttctttcgctttcc-5'
```

FIGURE 37

FIGURE 38

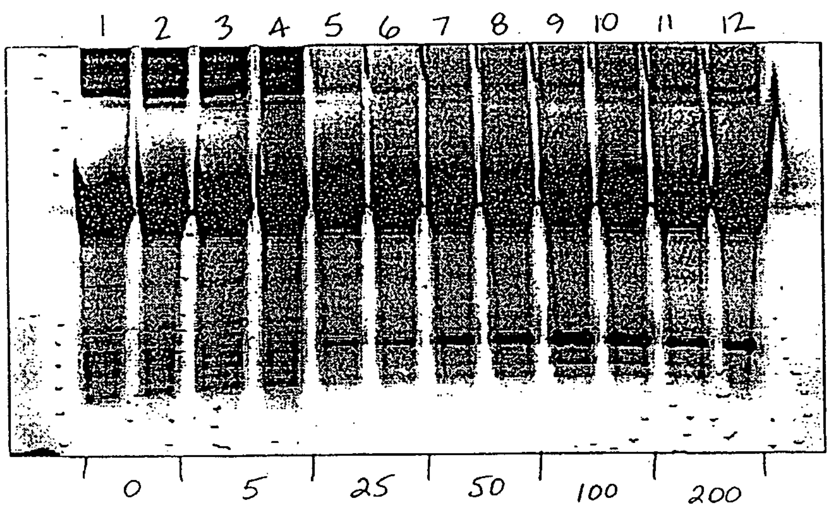


FIGURE 39

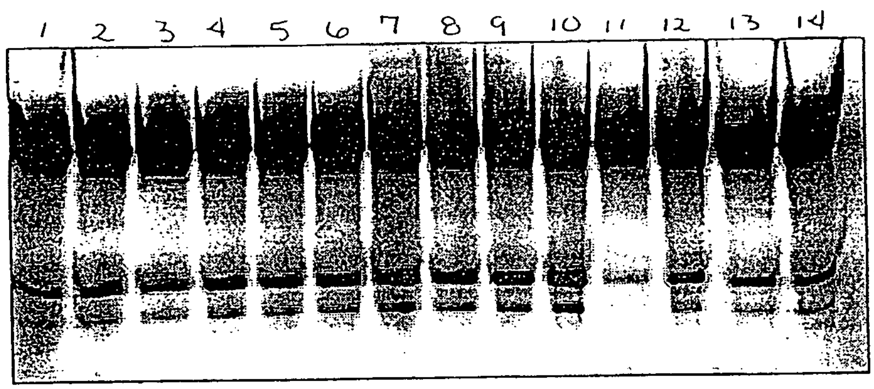


FIGURE 40

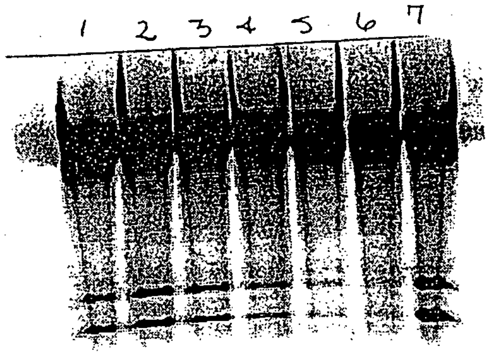


FIGURE 41

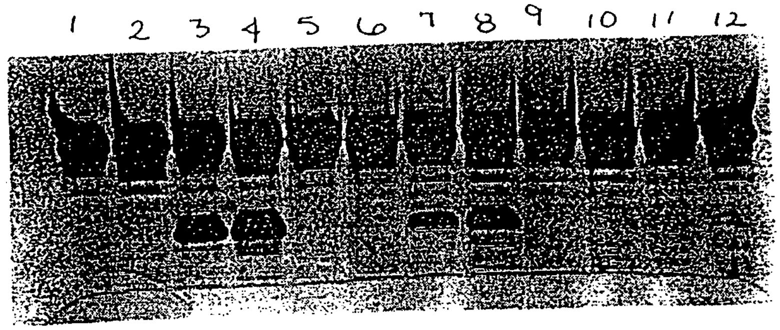


FIGURE 42A

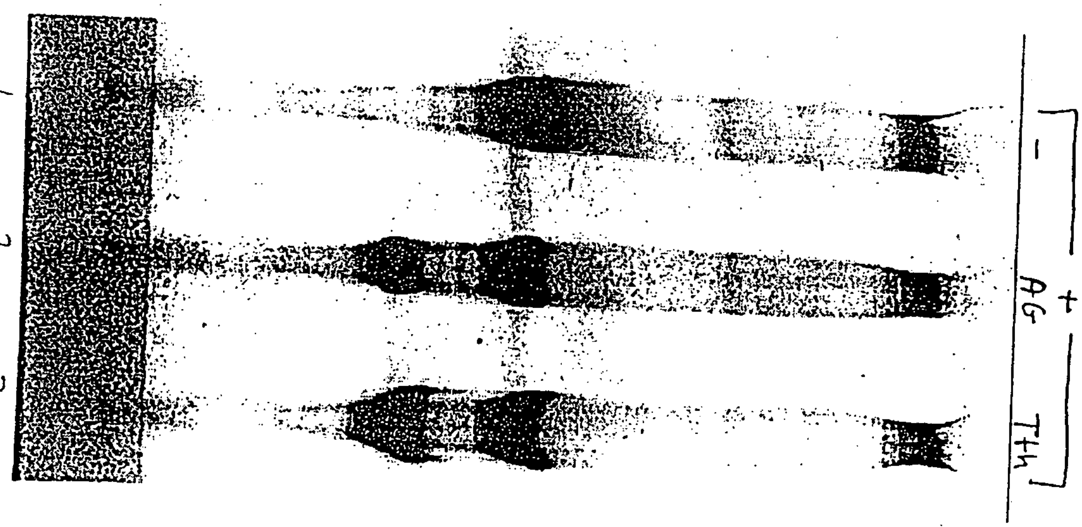
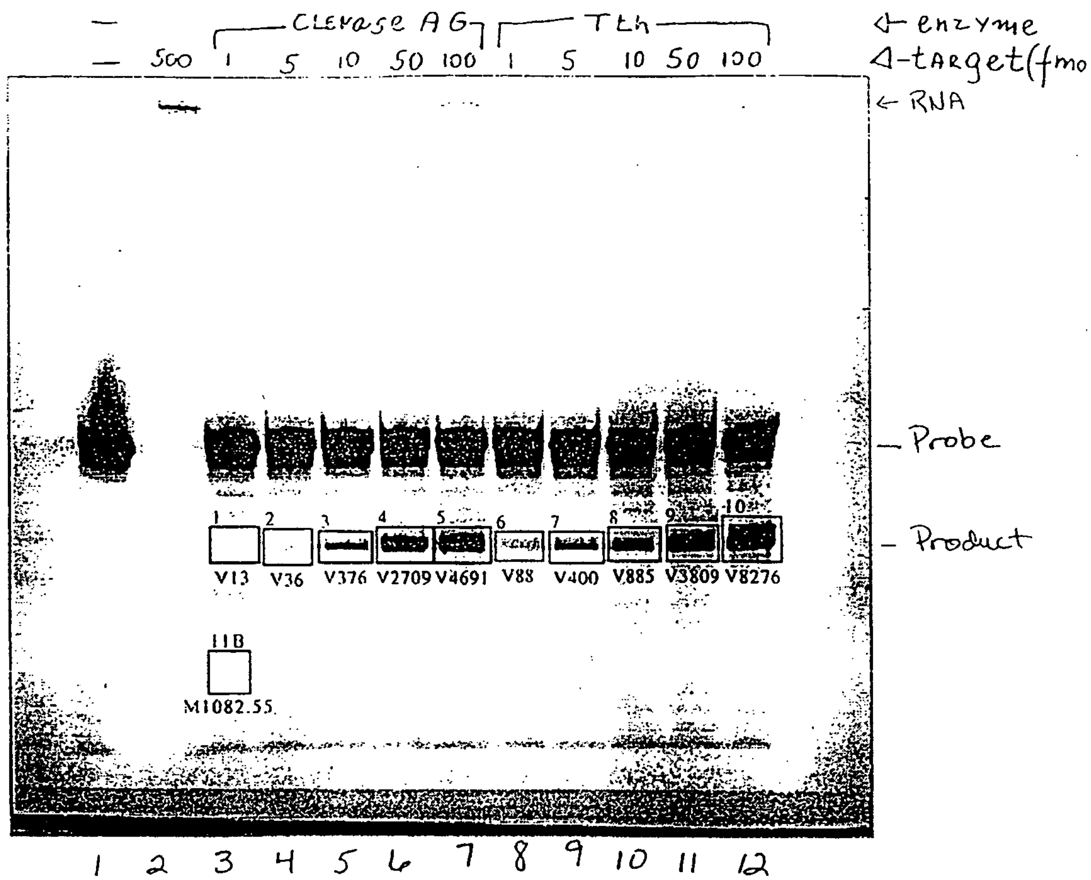


FIGURE 42B



Cleavase AG
Tth
← enzyme
← target(fmo
500
1
5
10
50
100
1
5
10
50
100
← RNA
— Probe
— Product
V13
V36
V376
V2709
V4691
V88
V400
V885
V3809
V8276
11B
M1082.55
1
2
3
4
5
6
7
8
9
10
11
12

FIGURE 43

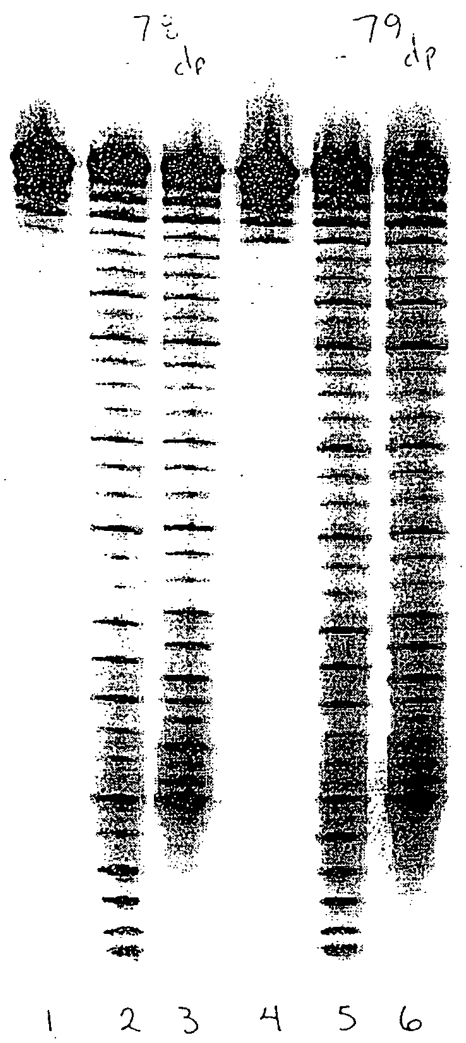
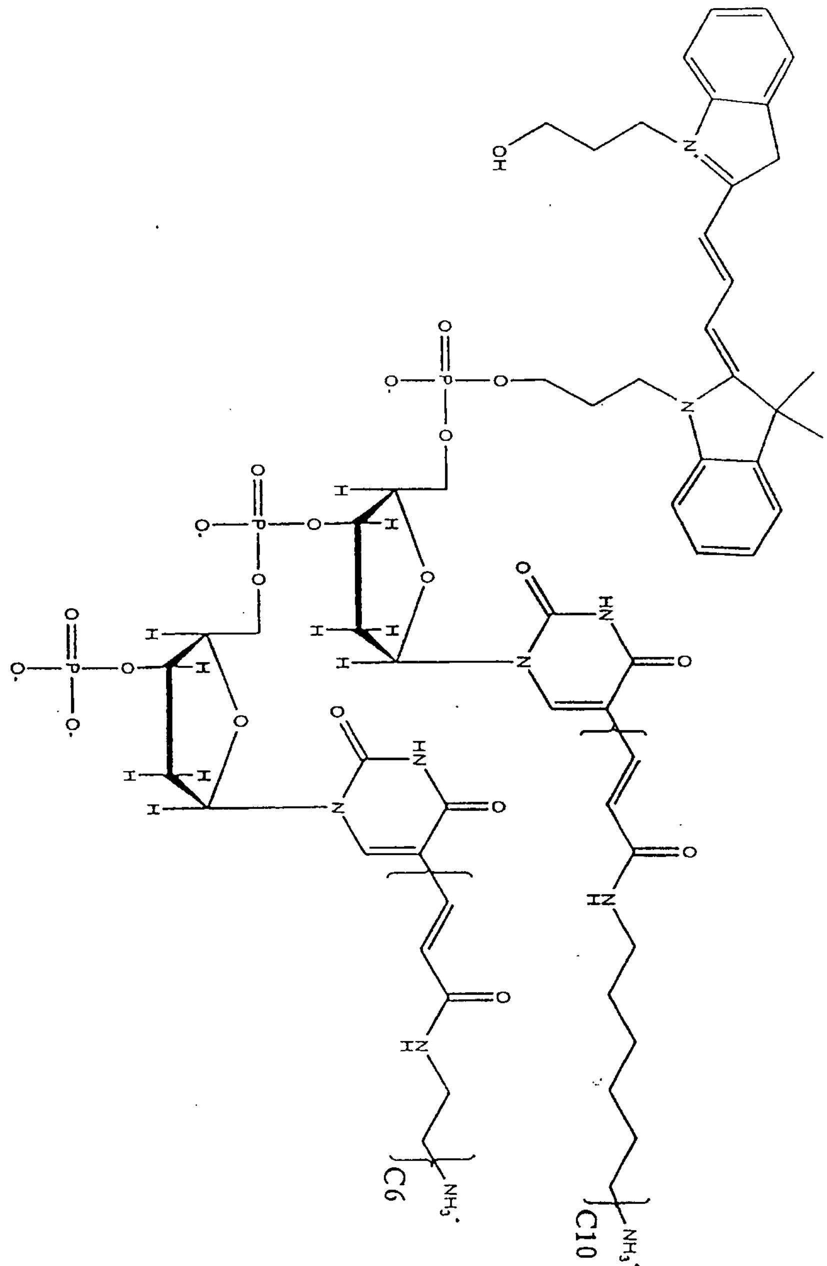
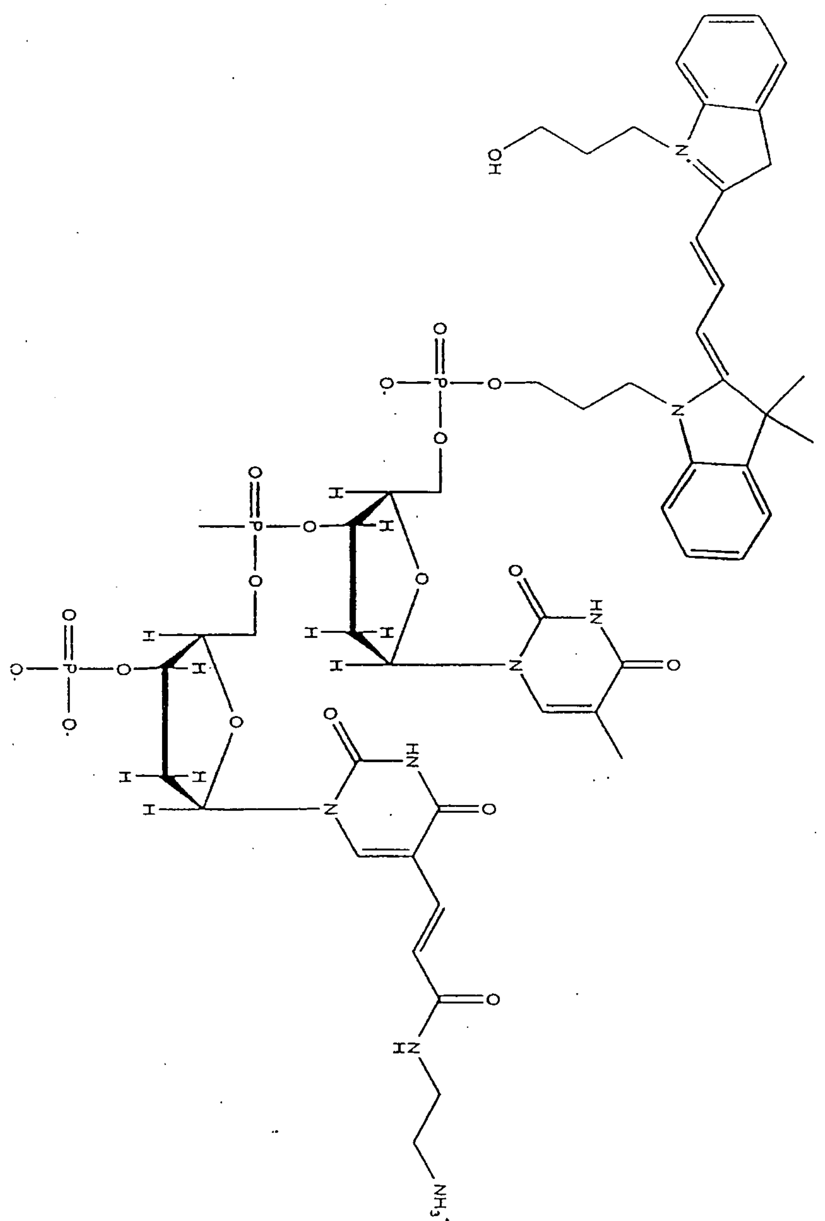


FIGURE 44

FIGURE 45

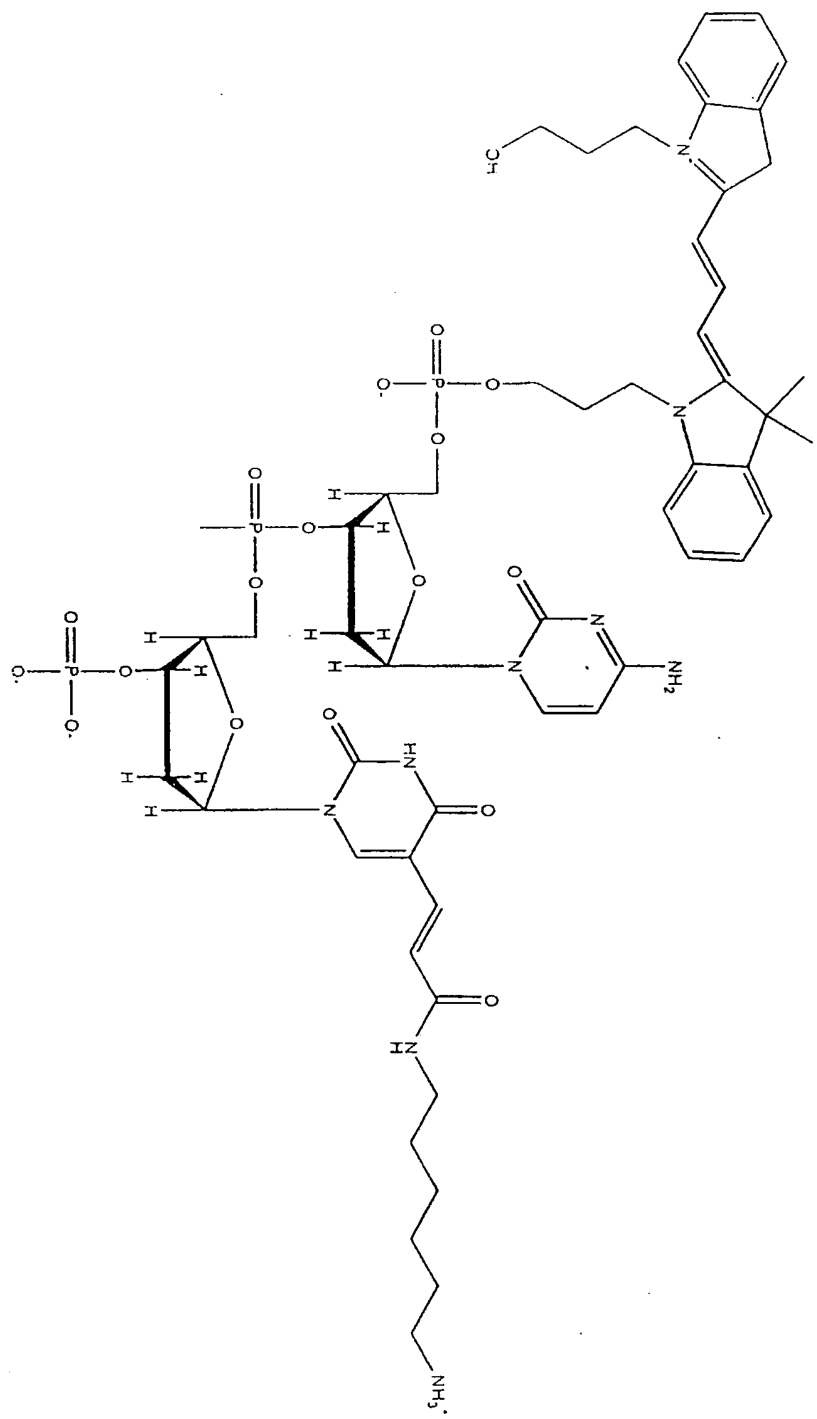
70 (C10 aminoT's)
74 (C6 amino T's)





75

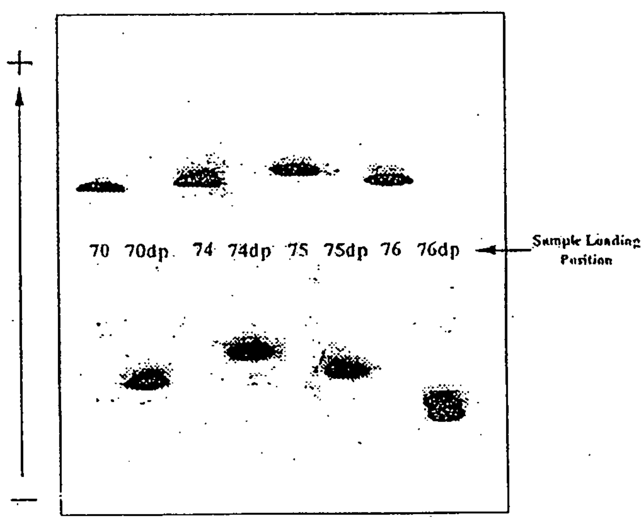
FIGURE 46



76

FIGURE 47

FIGURE 48



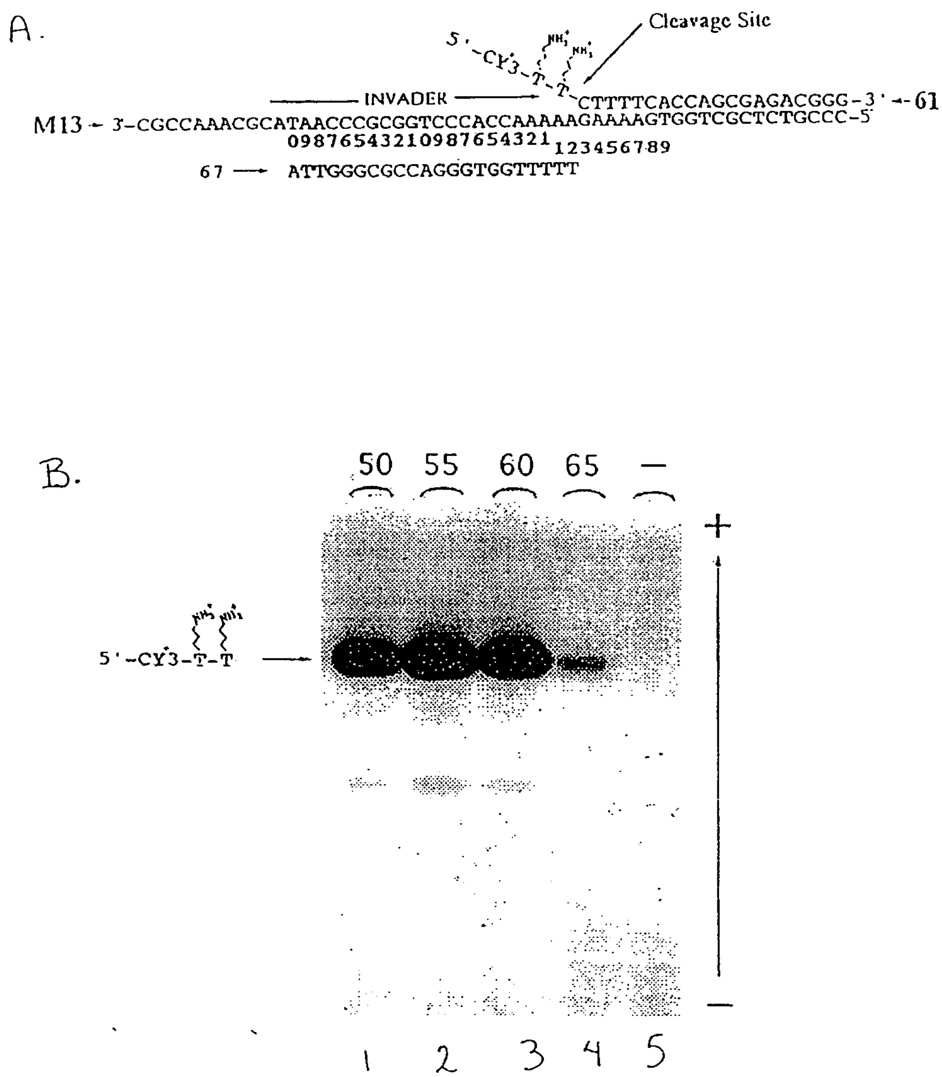
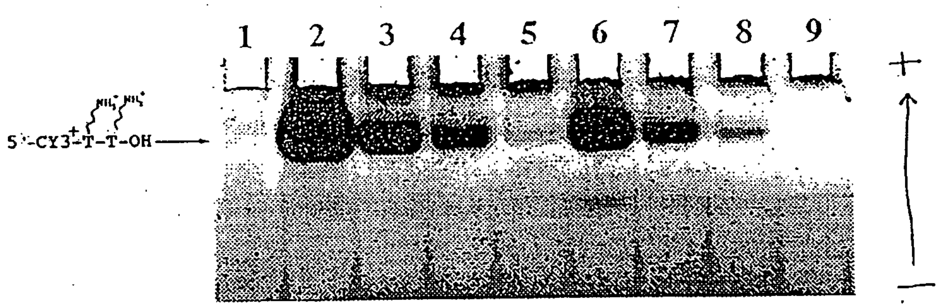

A.
Cleavage Site
5'-CY3-T-T
INVADER
CTTTTCACCAGCGAGACGGG-3' ← 61
M13 ← 3-CGCCAAACGCATAACCCGCGGTCCCACCAAAAAGAAAAGTGGTCGCTCTGCCC-5'
09876543210987654321 123456789
67 → ATTGGGCGCCAGGGTGGTTTTT
B.
50 55 60 65 —
+
5'-CY3-T-T
1 2 3 4 5

FIGURE 49

FIGURE 50



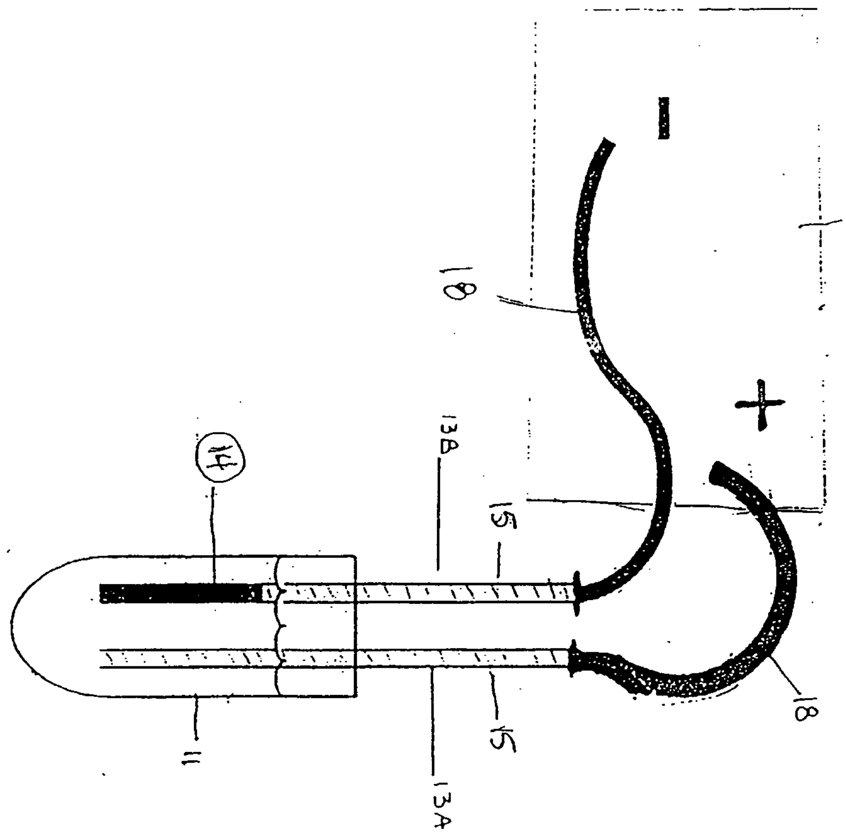


FIGURE 51

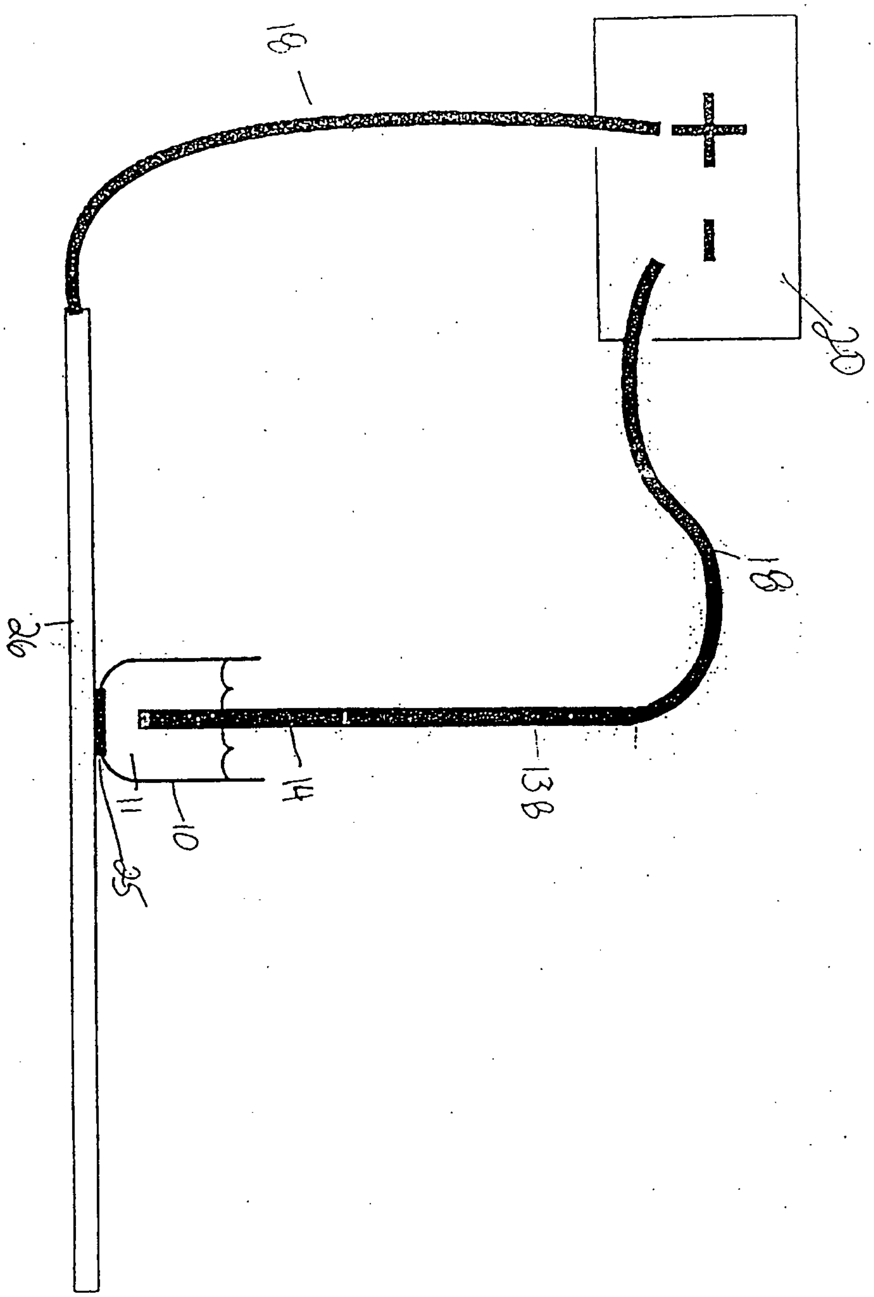
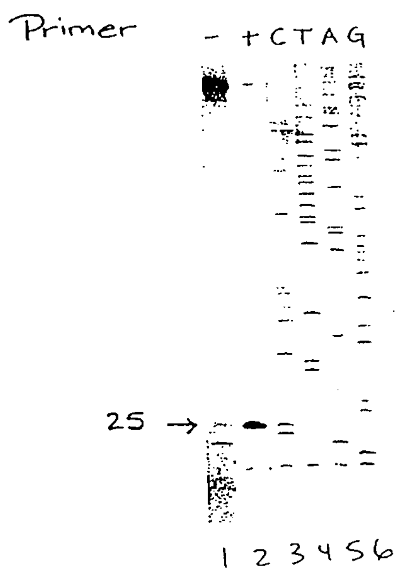


FIGURE 52

FIGURE 53



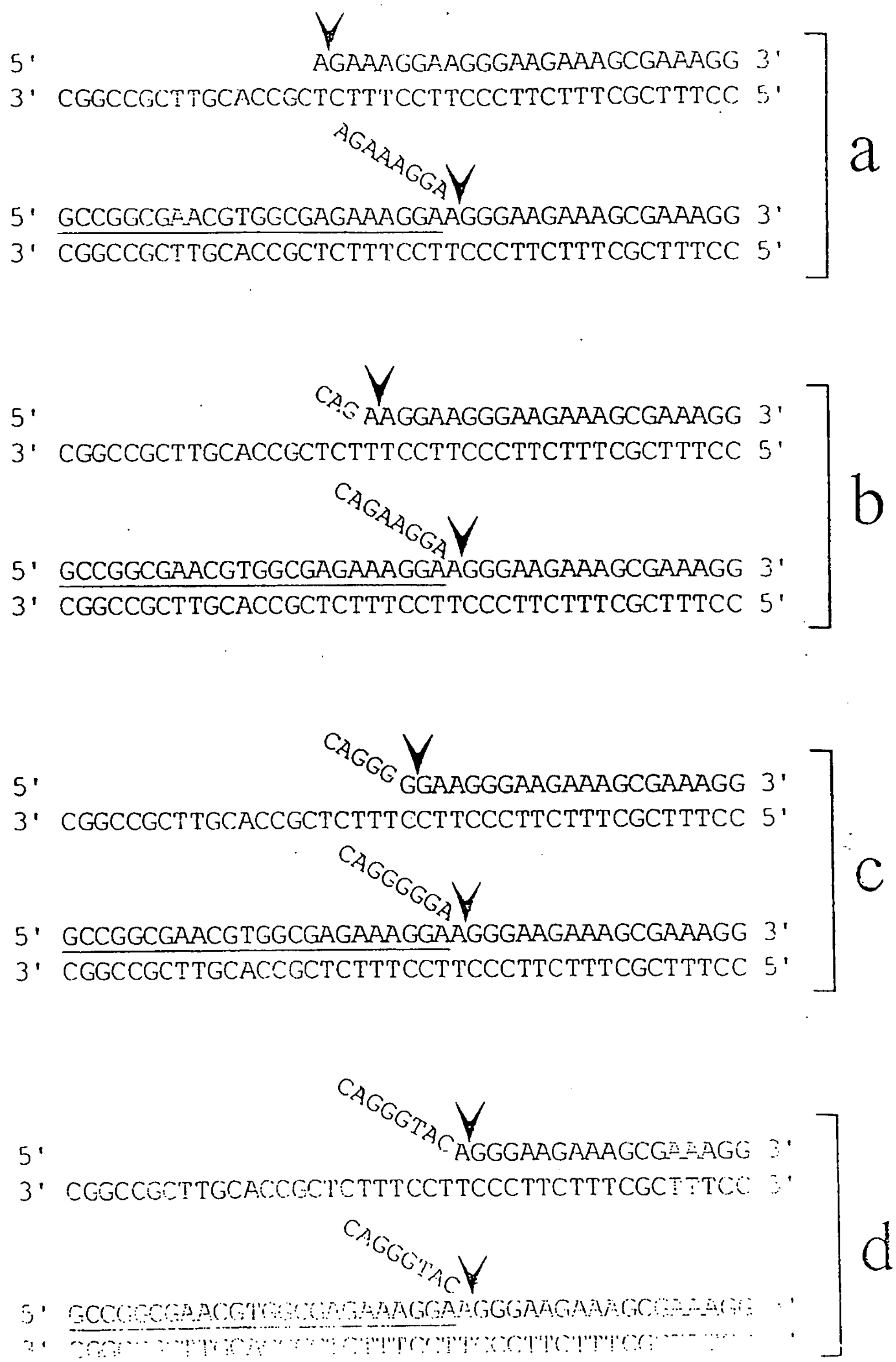
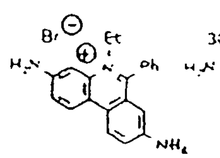
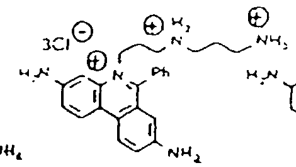

5' AGAAAGGAAGGGAAGAAAGCGAAAGG 3'
3' CGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCC 5'
AGAAAGGA
5' GCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGG 3'
3' CGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCC 5'
a
CAG
5' AAGGAAGGGAAGAAAGCGAAAGG 3'
3' CGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCC 5'
CAGAAAGGA
5' GCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGG 3'
3' CGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCC 5'
b
CAGGG
5' GGAAGGGAAGAAAGCGAAAGG 3'
3' CGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCC 5'
CAGGGGGA
5' GCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGG 3'
3' CGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCC 5'
c
CAGGGTAC
5' AGGGAAGAAAGCGAAAGG 3'
3' CGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTTTCC 5'
CAGGGTAC
5' GCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGG
d

FIGURE 54

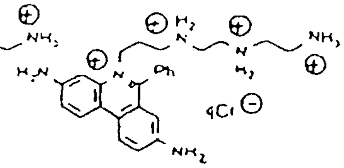
FIGURE 55



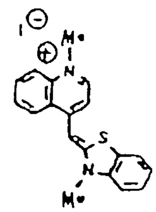
Ethidium Bromide



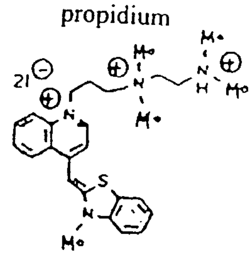
(1,3-propanediamino)-propidium



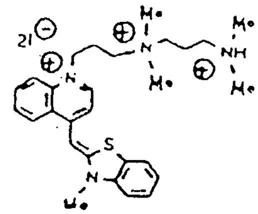
(diethylenetriamino)-propidium



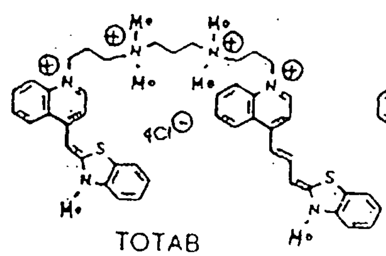
Thiazole Orange



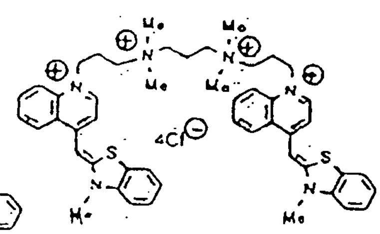
(N-N'-tetramethyl-1,2-ethanediamino)-propyl thiazole orange



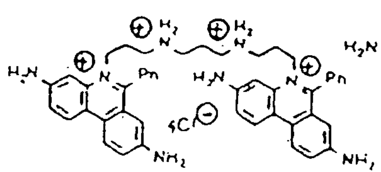
(N-N'-tetramethyl-1,3-propanediamino)-propyl thiazole orange



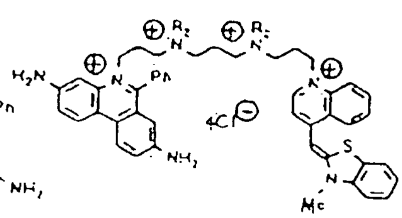
TOTAB



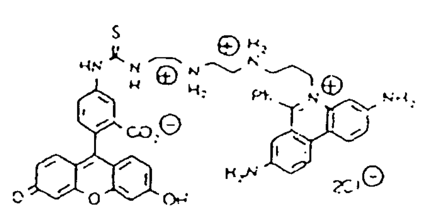
TOTO



EthD



TOED1 (R = H)
TOED2 (R = CH_3)



FED

FIGURE 56

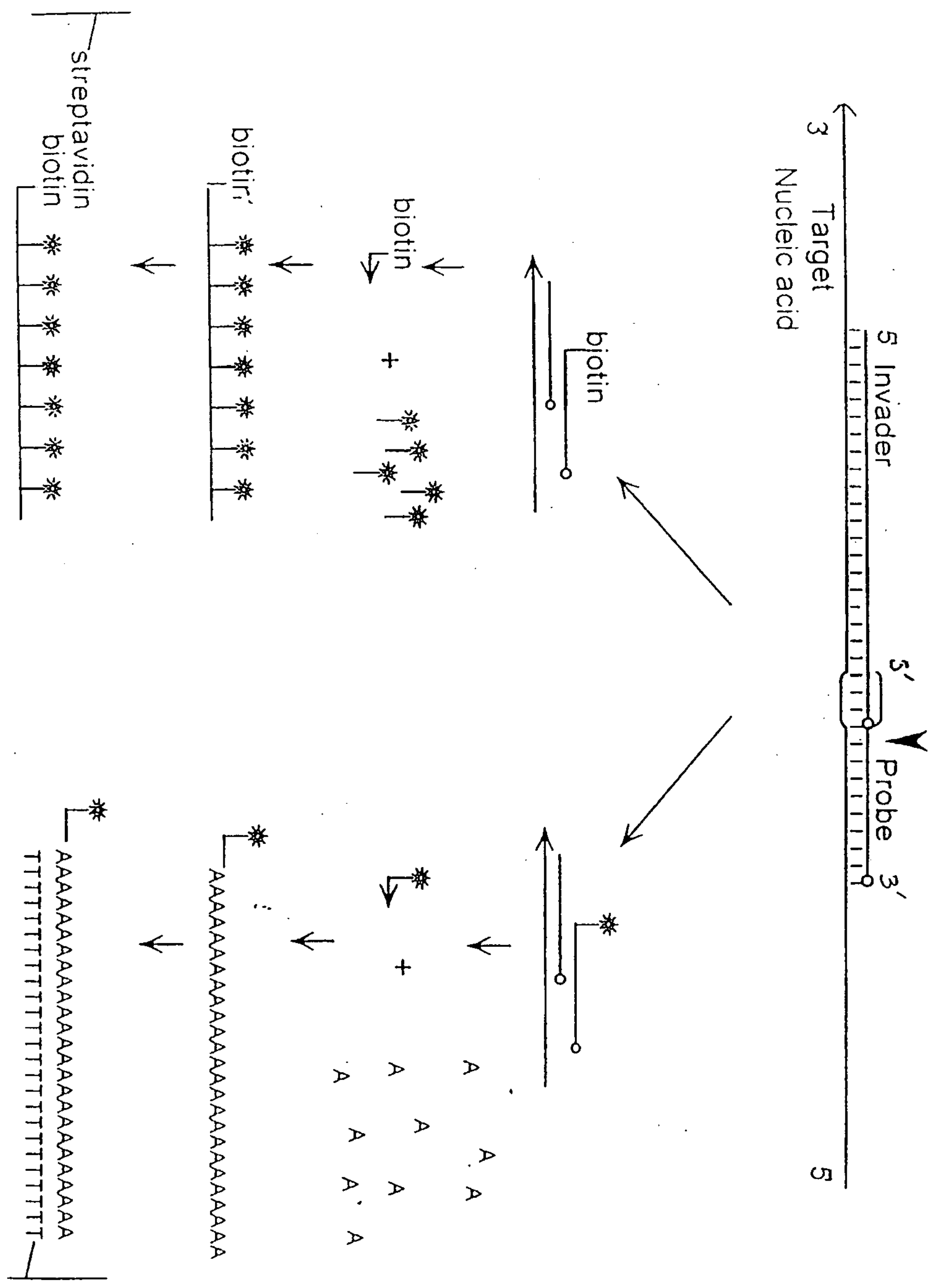

3
5
Target
Nucleic acid
5 Invader
3'
Probe
3'
biotin
biotin
+
+
biotin
streptavidin
biotin
AAAAAAAAAAAAAAAAAAAAAAAA
TTTTTTTTTTTTTTTTTTTTTTTT

FIGURE 57

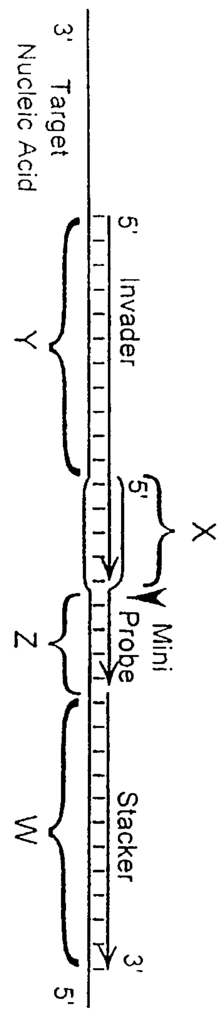


FIGURE 58

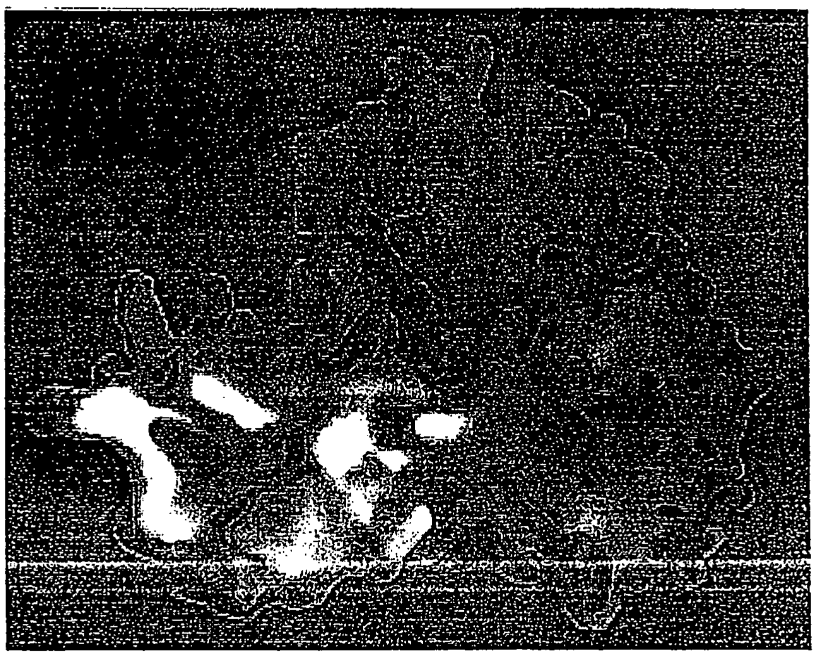


FIGURE 59A

```
                 20        30        40        50        60        70
------FGDFIPK--NIISFEDLKGKKVAIDGMNALYQFLTSIRLRDGSPLRNRKGEITSAYNGVFY MJAFEN1.PRO
------IGEIIPR--KEIELENLYGKKIAIDALNAIYQFLSTIRQKDGTPLMDSKGRITSHLSGLFY PFUFEN1.PRO
      IADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVRQ-GGDVLQNEEGETTSHLMGMFY HUMFEN1.PRO
      IADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVRQ-GGDVLQNEEGETTS-LMGMFY MUSFEN1.PRO
      ISEHVPSAIRKSDIKSFFGRKVAIDASMSLYQFLIAVRQQDGGQLTNEAGETTSHLMGMFY YST510.PRO
      IAG---PTARPVRLESLEDKRMAVDASIWIYQFLKAVRDQEGNAVKN------SHITGFFR YSTRAD2.PRO
      LE----PVKRPVKLETLVNKRLAIDASIWIYQFLKAVRDKEGNQLKS------SHVVGFFR SPORAD13.PRO
      LE----CSGRQVSPEALEGKILAVDISIWLNQALKGVRDRHGNSIEN------PHLLTLFH HUMXPG.PRO
      LE----CSGHRVSPEALEGKVLAVDISIWLNQALKGVRDSHGNVIEN------AHLLTLFH MUSXPG.PRO
      LE----CSGRPINPGTLEGKILAVDISIWLNQAVKGARDRQGNAIQN------AHLLTLFH XENXPG.PRO
      ANHVV----RKVPNETMRDKTLSIDGHIWLYESLKGCEAHHQQT-------PNSYLVTFFT CELRAD2 PRO

       80        90       100       110       120       130       140
      ITPIWVFDGEPPKLKEKTRKVRREMKEKAELKMKEAIKK----EDFEEAAKYAKRVSYLTP MJAFEN1 PRO
      IKPVYVFDGEPPEFKKKELEKRREAREEAEEKWREALEK----GEIEEARKYAQRATRVNE PFUFEN1 PRO
      IKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQAQAA----GAEQEVEKFTKRLVKVTK HUMFEN1.PRO
      IKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQAQEA----GMEEEVEKFTKRLVKVTK MUSFEN1.PRO
      IKPCYVFDGKPPDLKSHELTKRSSRRVETEKKLA---EA----TTELEKMKQERRLVKVSK YST510.PRO
      IRPVFVFDGGVPVLKRETIRQRKERRQGKRESAKSTARKLLALQLQNGSNDNKRDSDEVTM YSTRAD2.PRO
      IKPVFVFDGGAPSLKRQTIQKRQARRLDREENATVTANKLLALQMRHQAMLLKRDADEVTQ SPORAD13.PRO
      IRPIFVFDGDAPLLKKQTLVKRRQRKDLASSDSRKTTEKLLKTFLKRQAIKTERIAATVTG HUMXPG.PRO
      IRPIFVFDGDAPLLKKQTLAKRRQRKDSASIDSRKTTEKLLKTFLKRQALKTDRIAASVTG MUSXPG PRO
      IRPIFVFDGEAPLLKRQTLAKRRQRTDKASNDARKTNEKLLRTFLKRQAIKAERIAATVTG XENXPG.PRO
      IPIVVFDNINASSSAHESKDQNEFVPRKRRSFGDSPFTNLV------------------- CELRAD2 PRO
```

FIGURE 59B

```
        150       160       170       180       190       200       210
 YVENCKYLLSLMGIPYVEAPSEGEAQASYMAKKGDVWAVVSQDYDALLYGAPRVVRNLTTTKEM-----  MJAFEN1.PRO
 TLIEDAKKLLELMGIPIVQAPSEGEAQAAYMAAKGSVYASASQDYDSLLFGAPRLVRNLTITGKRKLPGK  PFUFEN1.PRO
 AHMDECKHLLSLMGIPYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFGSPVLMRHLTASEAKKLPIQ  HUMFEN1.PRO
 AHMDECKHLLSLMGIPYLDAPSEAEASCAALAKAGKVYAAATEDMDCLTFGSPVLMRHLTASEAKKLPIQ  MUSFEN1.PRO
 THEEAQKLLGLMGIPYIIAPTEAEAQCAELAKKGKVYAAASEDMDTLCYRTPFLLRHLTFSEAKKEPIH  YST510.PRO
 TKEVQELLSRFGIPYITAPMEAEAQCAELLQLNLVDGIITDDSDVFLFGGTKIYKNMFHEKNY---VE  YSTRAD2.PRO
 NECQELLRLFGLPYIVAPQEAEAQCSKLLELKLVDGIVTDDSDVFLFGGTRVYRNMFNQNKF---VE  SPORAD13.PRO
 TFLESQELLRLFGIPYIQAPMEAEAQCAILDLTDQTSGTITDDSDIWLFGARHVYRNFFNKNKF---VE  HUMXPG.PRO
 MFLESQELLRLFGVPYIQAPMEAEAQCAVLDLSDQTSGTITDDSDIWLFGARHVYKNFFNKNKF---VE  MUSXPG.PRO
 MQLESQELLQLFGIPYIVAPMEAEAQCAILDLTDQTSGTITDDSDIWLFGARHVYKNFFSQNKH---VE  XENXPG.PRO
 TNALLTELGIKVIIAPGDGEAQCARLEDLGVTSGCITTQFDYFLFGGKNLYRFDFTAGT------  CELRAD2.PRO

        220       230       240       250       260       270       280
 -------PELIELNEVLEDLRISLDQLIDIAIFMGTDYNPGGV--K--GIGFKRAYELVRSGVAK--DV  MJAFEN1.PRO
 VE-IKPELIILEEVLKELKLTREKLIELAILVGTDYNPGGI--K--GIGLKKALEIVRHSKDPLAKF  PFUFEN1.PRO
 EFHLSRILQELGLNQEQFVDLCILLGSDYCESIRGIGPKRAVDLIQK--HKSIEEIVRRLDPN-----KY  HUMFEN1.PRO
 EFHLSRVLQELGLNQEQFVDLCILLGSDYCESIRGIGAKRAVDLIQK--HKSIEEIVRRLDPS-----KY  MUSFEN1.PRO
 TELVLRGLDLTIEQFVDLCIMLGCDYCESIRGVGPVTALKLIKT--HGSIEKIVEFIESGESNNTKW  YST510.PRO
 TDAESILKLLGLDRKNMIELAQLLGSDYTNGLKGMGPVSSIEVIAEF---GNLKNFKDWYNNGQFDKRK  YSTRAD2.PRO
 LMDDMKREFNVNQMDLIKLAHLLGSDYTMGLSRVGPVLALEILHEFPGDTGLFEFKKWFQRLSTGHAS  SPORAD13.PRO
 CYVDFHNQLGLDRNKLINLAYLLGSDYTEGIPTVGCVTAMEILNEFPGHGLEPLLKFSEWWHEAQKNP  HUMXPG.PRO
 FFSQLGLDRNKLINLAYLLGSDYTEGIPTVGCVTAMEILNEFPGRGLDPLLKFSEWWHEAQNNK  MUSXPG.PRO
 MQLDRSKLINLAYLLGSDYIIGIPTVGYVSAMEILNEFPGQGLEPLVKFKEWWSEAQKDK  XENXPG.PRO
 -----------------------------------SSTACLHDIMHLSLGRMFM------------  CELRAD2.PRO
```

FIGURE 59C

```
          290       300       310       320       330       340       350
...IKRIFKEPKV-------------------TD--NYSLSLKLPDKEGIIKFLVDENDFNYD  MJAFEN1.PRO
...AIKEFFLNPPV-------------------TD--NYNLVWRDPDEEGILKFLCDEHDFSEE  PFUFEN1.PRO
...LAHOLFLEPEV-------------------LDPESVELKWSEPNEEELIKFMCGEKOFSEE  HUMFEN1 PRO
...EAOOLFLEPEV-------------------VDPESVELKWSEPNEEELVKFMCGEKOFSEE  MUSFEN1.PRO
...KOARMLFLDPEV-------------------IDGNEINLKWSPPKEKELIEYLCDDKKFSEE  YST510.PRO
...ENKFEKDLRKKLVNNEIILDDDFPSVMVYDAYMRPEVDHDTTPFVWGVPDLDMLRSFMKTOLGWPHE  YSTRAD2.PRO
...DVNTPVKKRINKLVGK-IILPSEFPNPLVDEAYLHPAVDDSKOSFOWGIPDLDELROFLMATVGWSKO  SPORAD13.PRO
...DTKVKKKL--RTLOLTPGFPNPAVAEAYLKPVVDDSKGSFLWGKPDLDKIREFCORYFGWNRT  HUMXPG PRO
...DTKVKKKL--RKLOLTPGFPNPAVADAYLRPVVDDSRGSFLWGKPDVDKIREFCORYFGWNRM  MUSXPG.PRO
...DTKVKKKL--RLLDLOOSFPNPAVASAYLKPVVDESKSAFSWGRPDLEOIREFCESRFGWYRL  XENXPG.PRO
----EKKVSRPHLISTAILLGCDYFORGVONIGIVSVFD-ILGEFGDDGNEEIDPHVILDRFASYVRE  CELRAD2.PRO

          360       370       380       390       400       410       420
...KHVDKLYNLIA---------------------------------------------------  MJAFEN1.PRO
...GLERLKKAI----------------------------------------------------  PFUFEN1 PRO
...SKSROGS-TOGRLDDFFKVT-----------------------------------------  HUMFEN1.PRO
...GVKRLSKSROGS-TOGRLDDFFKVT-----------------------------------------  MUSFEN1.PRO
...SRLKKGLKSG-IOGRLDGFFOVV-----------------------------------------  YST510 PRO
...IE..IPLIRDVNKRKK---------------------------------------KGKO  YSTRAD2.PRO
...FVLLPVIODMHKKOF--------------------------------------VGTO  SPORAD13 PRC
...ESLFPVLKOLDAOOTOLRIDSFFRLAOOEKEDAKRIKSORLNRAVTCMLRKEKEAAASEIEAVSVAM  HUMXPG PRO
...ESLYPVLKHLNAHOTOLRIDSFFRLAOOEKODAKLIKSHRLSRAVTCMLRKEREEKAPELTKVTEAM  MUSXPG PRO
...KOLNAOOTOLRIDSFFRLEOHEAAG---LKSORLRRAVTCMKRKERDVEAEEVEAAVAVM  XENXPG PRO
...PKLRLRRKKYNFPVGFPNCDAVHNAITMYLRPPVSSEIPKIIPR-----AANFOOVAEIM  CELRAD2 PRO
```

FIGURE 59D

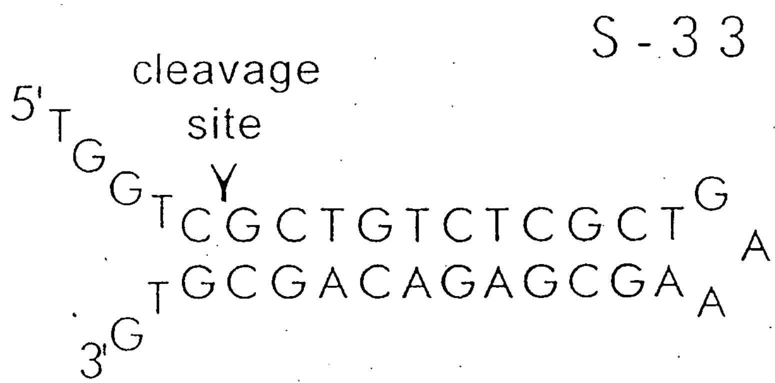
```
        430       440       450       460       470       480       490
------------------------------------------------------------------------ MJAFEN1.PRO
------------------------------------------------------------------------ PFUFEN1.PRO
--------------------------------------------------------------------GSLS HUMFEN1.PRO
--------------------------------------------------------------------GSLS MUSFEN1.PRO
--------------------------------------------------------------------PK-T YST510.PRO
EFF--------------------------------------------------------------------- YSTRAD2.PRO
FDGNINVYAPRVAYHFKSKRLENALSSFKNQISNQSPMSEEIQADADAFGESKGSDELQSRIL SPORAD13.PRO
DKAKRKTQKRGITNTLEESSSLKRKRLSDSKRKNTCGGFLGETCLSESSDGSSSEHAESSSLM HUMXPG.PRO
IFELLDDAKGKTQKRELPYK-------KETSVPKRRRPSGNGGFLGDPYCSESPQESSCEDGEGSSVM MUSXPG.PRO
EDTNQRKGQKTNTKS-------QGTKRRKPTECSQEDQDPGGGFIGIELKTLSSKAYSSD-------- XENXPG.PRO
EDGMPATRTQKELALSIRRKVHLTTTVAQTRIPDFFAATKSKNFTPIVEPCESLEDYISANN-----T CELRAD2.PRO

        500       510       520       530       540       550       560
----------------------------------------------------------------NKTKQKTL MJAFEN1.PRO
----------------------------------------------------------------KSGKQSTL PFUFEN1.PRO
TKEPEPKGST------------------------------------------------KKKAKTGAAG HUMFEN1.PRO
AKEPEPKGPA------------------------------------------------KKKAKTGGAG MUSFEN1.PRO
PAQE--------------------------------------------------NKKLNKNKNK YST510.PRO
-------------------------PREYISGDKKLNTSKRISTATGKL----------------KK- YSTRAD2.PRO
SDSDSDSEDNFLASLTPKTNSSSISIENLPRKTKLSTSLL----------------KKP SPORAD13.PRO
PKTSASDSQNSVKEAPVKNGGATTSSSSDSDDDGGKEKMVLVTARSVFGKKRRKLRRARG HUMXPG.PRO
SKIGCSDVPDLVRDSPHGRQGCVSTSSSDSEDGEDKAKTVLVTARPVFGKKRRKLKSMK- MUSXPG.PRO
---DAEDLPSGLIDKQSQSGIVGRQKASNKVESSSSSDDEDRTVMVTAKPVFQGKKTKSKTMKE XENXPG.PRO
--RKPGIIQHHAKRQVPQRK---------------------------RSVKIRAFKPYPTDVI CELRAD2.PRO
```

FIGURE 59E

[illegible]AFKZ	MJAFEN1.PRO
[illegible]AFKR	PFUFEN1 PRO
[illegible]RGK	HUMFEN1 PRO
[illegible]RGK	MUSFEN1.PRO
[illegible]GRR	YST510.PRO
[illegible]RK[illegible]	YSTRAD2.PRO
[illegible]RRK	SPORAD13.PRO
[illegible]	HUMXPG PRO
[illegible]K[illegible]	MUSXPG.PRO
[illegible]RK	XENXPG.PRO
[illegible]DSD	CELRAD2 PRO

FIGURE 60


S-33
cleavage site
5' TGGT
CGCTGTCTCGCT
G A A
TGCGACAGAGCGA
3' G

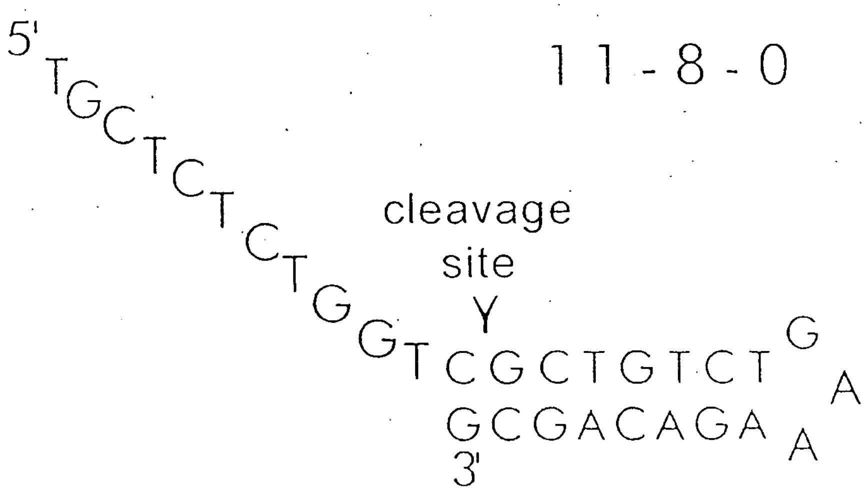

11-8-0
cleavage site
5' TGCTCTCTGGT
CGCTGTCT
G A A
GCGACAGA
3'

FIGURE 61

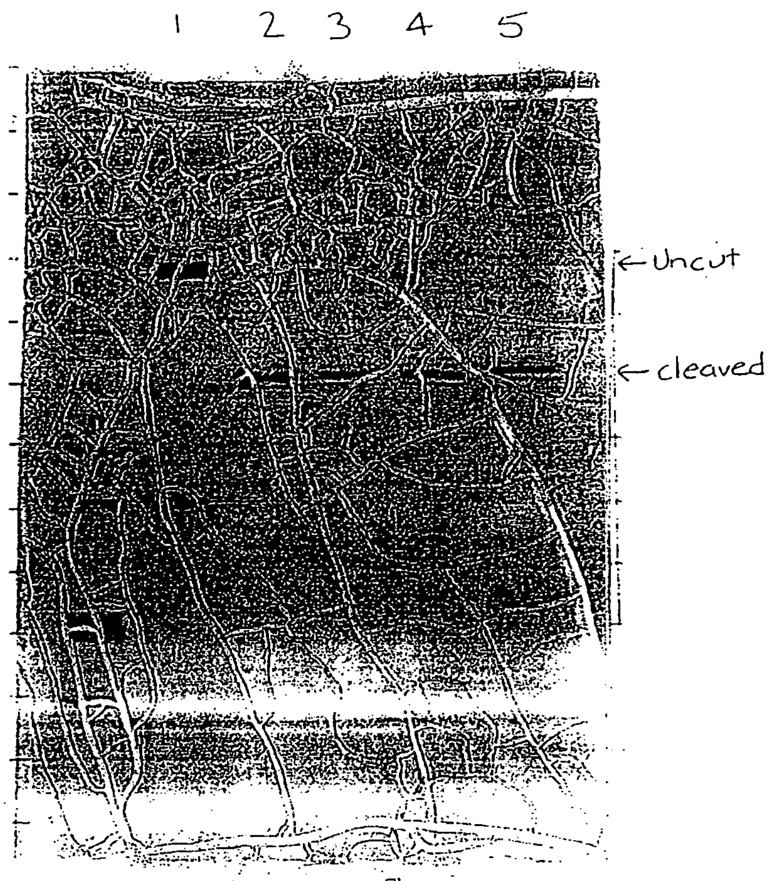


FIGURE 62

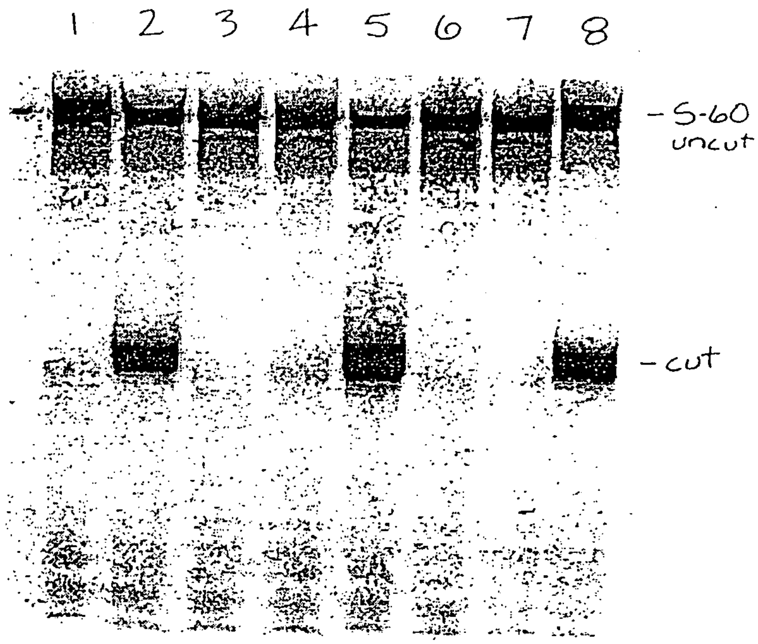
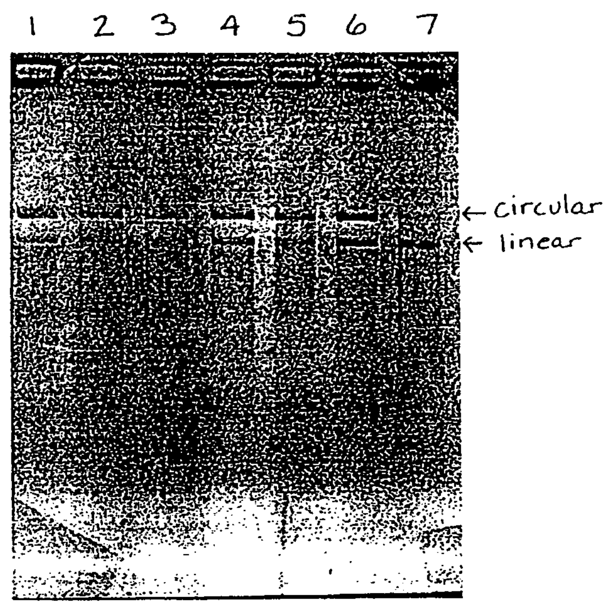


FIGURE 63



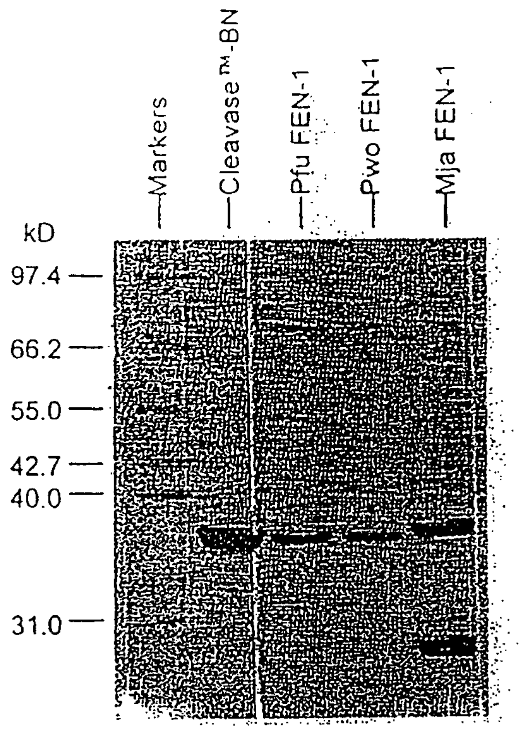

Markers
Cleavase™-BN
Pfu FEN-1
Pwo FEN-1
Mja FEN-1
kD
97.4
66.2
55.0
42.7
40.0
31.0

FIGURE 64

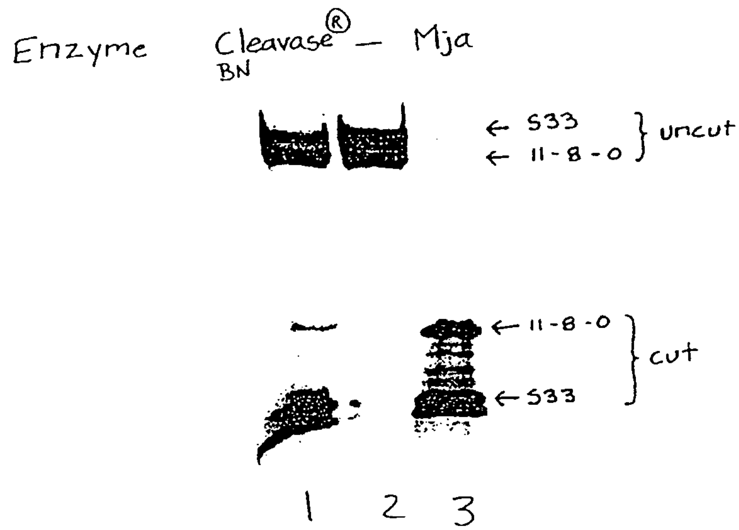
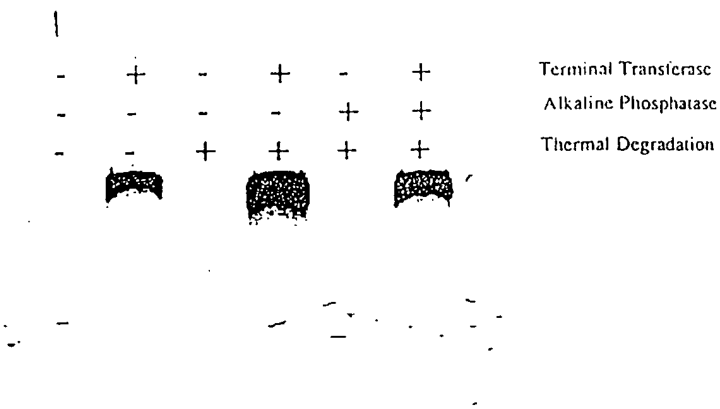


FIGURE 65



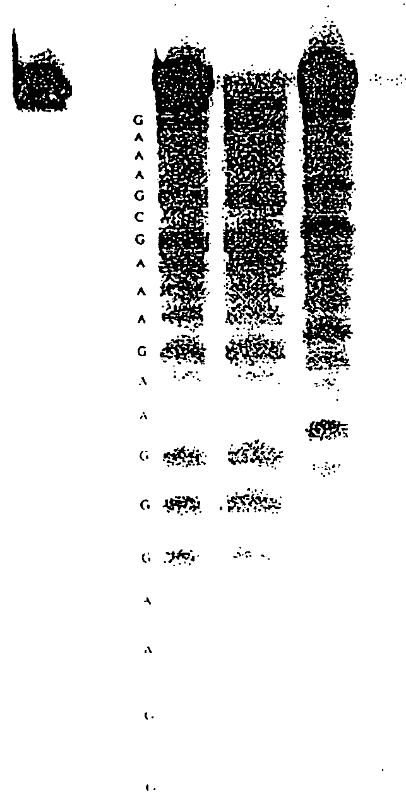


FIGURE 66

5'-nAGAAAggaaggga agaaagcgaaagG-3'

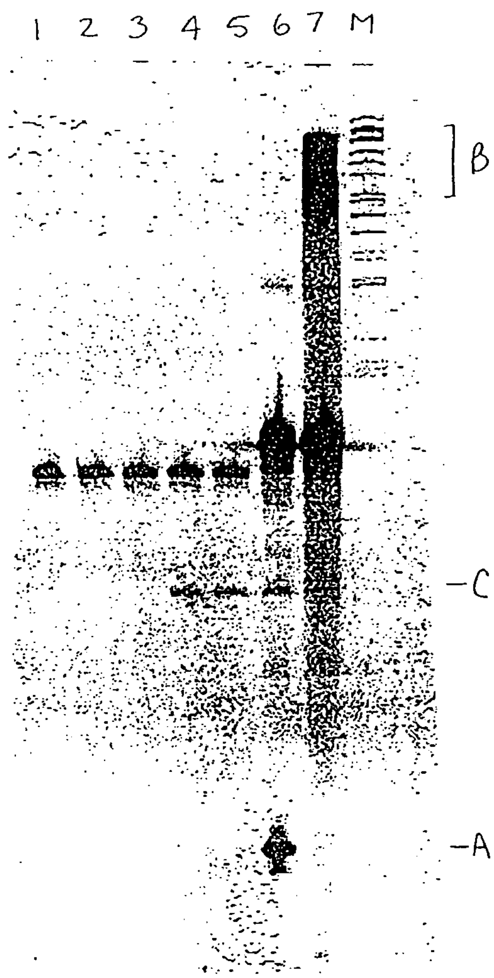


FIGURE 67

FIGURE 68

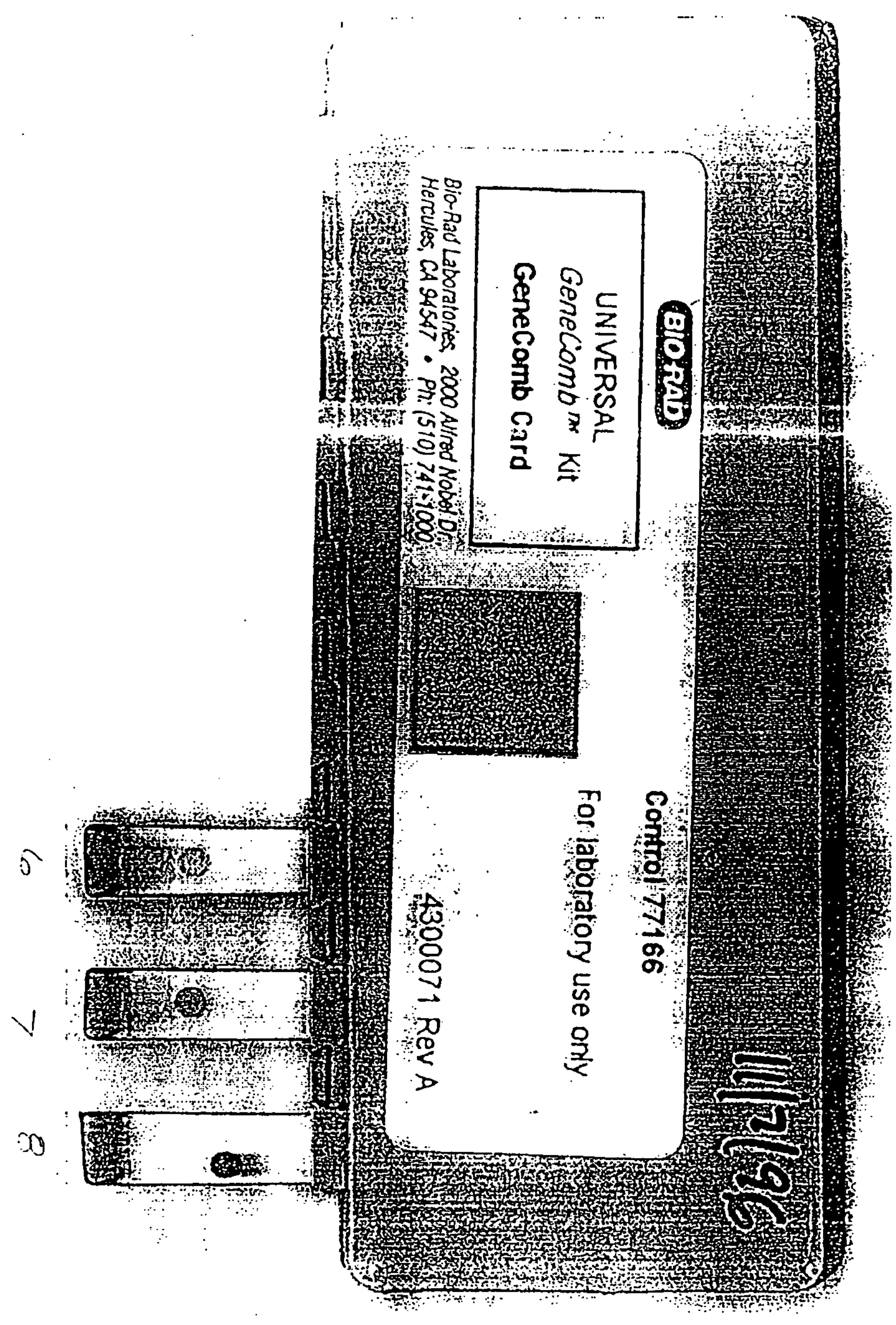

BIO-RAD
UNIVERSAL
GeneComb™ Kit
GeneComb Card
Bio-Rad Laboratories, 2000 Alfred Nobel Dr.
Hercules, CA 94547 • Ph: (510) 741-1000
Control 77166
For laboratory use only
4300071 Rev A
6
7
8

FIGURE 69

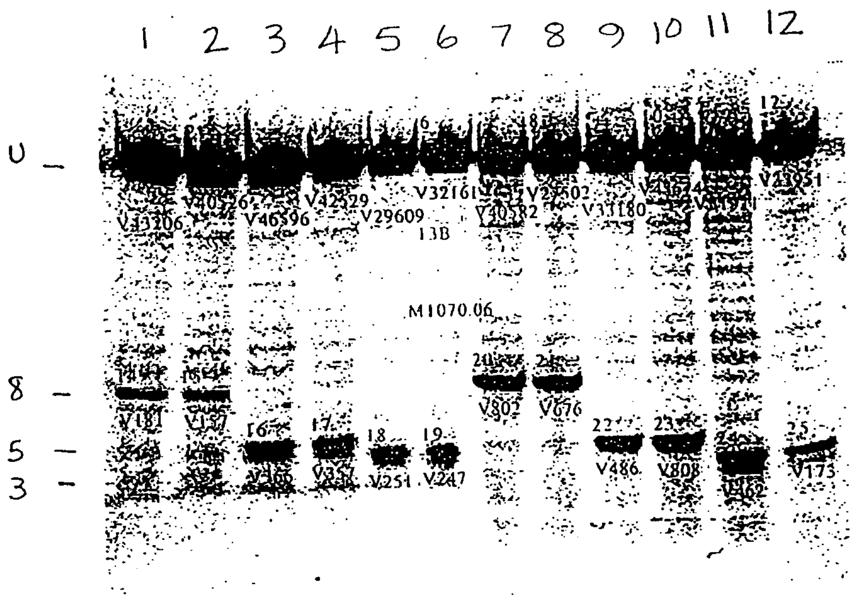

1 2 3 4 5 6 7 8 9 10 11 12
U
8
5
3
M1070.06
V32161
V29609
V46596
V33180
V486
V808
V802
V676
V251
V247
V173

FIGURE 70

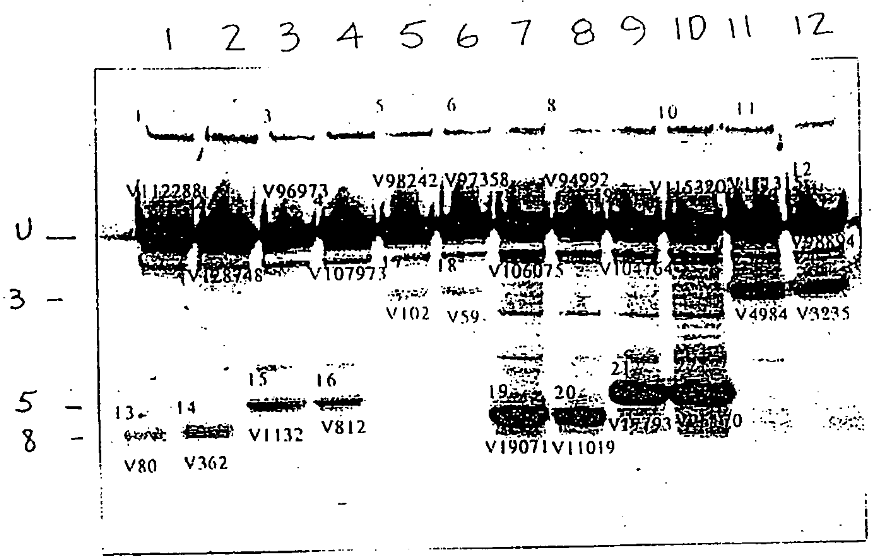
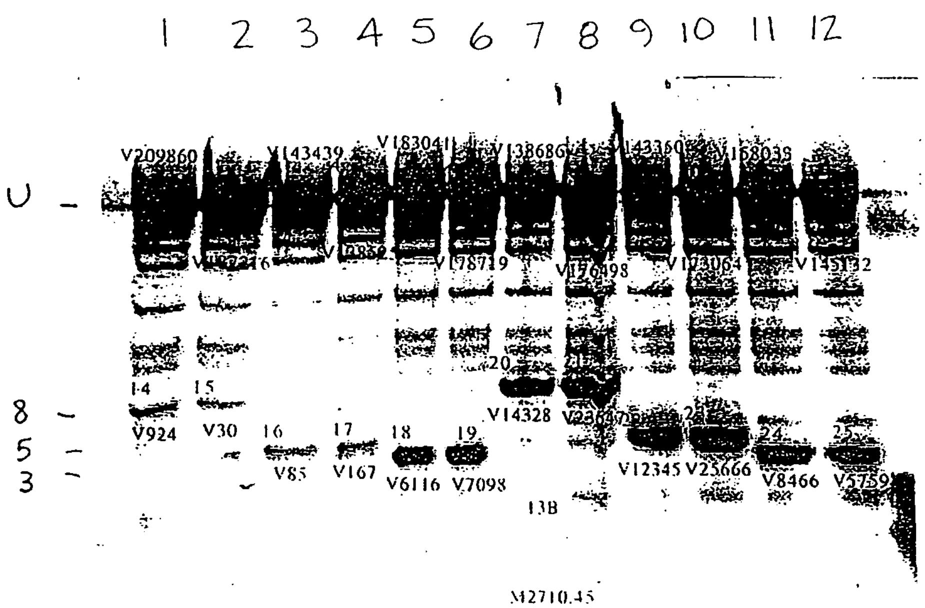

1 2 3 4 5 6 7 8 9 10 11 12
U
3
5
8
V112288
V96973
V98242
V97358
V94992
V115390
V128748
V107973
V106075
V104764
V93894
V102
V59
V4984
V3235
V80
V362
V1132
V812
V19071
V11019

FIGURE 71


1 2 3 4 5 6 7 8 9 10 11 12
U
8
5
3
V209860
V143439
V183041
V138686
V143360
V168039
V176498
V178719
V145132
V924
V30
V85
V167
V6116
V7098
V14328
V12345
V25666
V8466
14
15
16
17
18
19
20
24
25
M2710.45

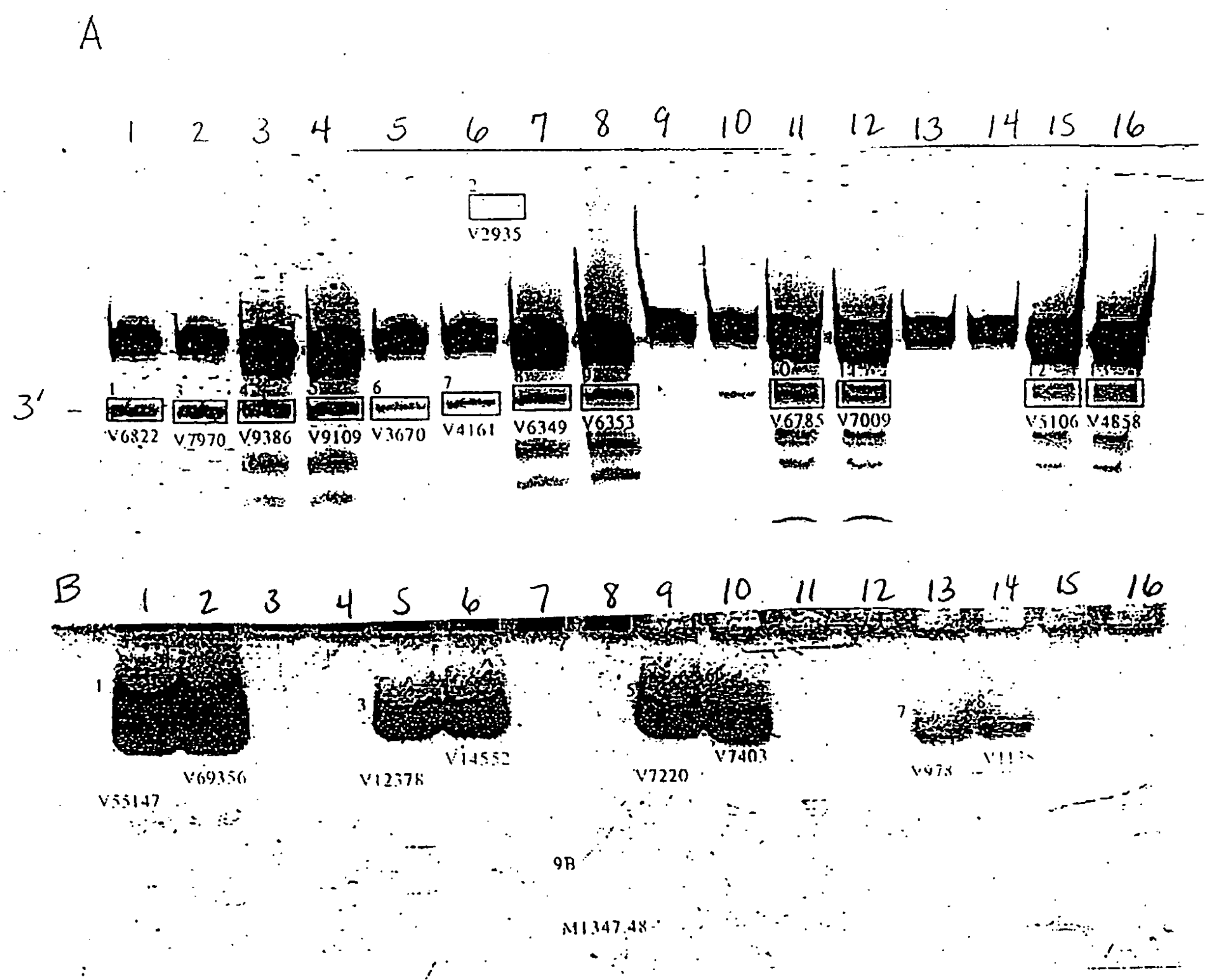
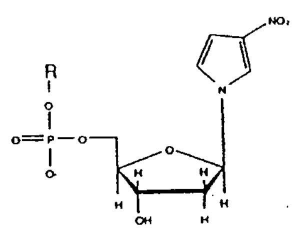
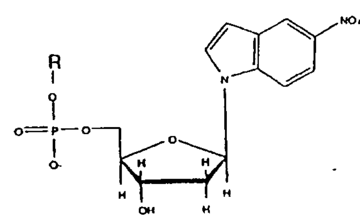

A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
V2935
3'
V6822
V7970
V9386
V9109
V3670
V4161
V6349
V6353
V6785
V7009
V5106
V4858
B
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
V55147
V69356
V12378
V14552
V7220
V7403
9B
M1347.48

FIGURE 72

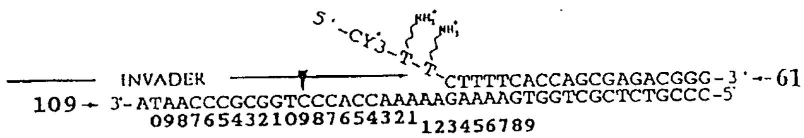
FIGURE 73



3-nitropyrrole



5-nitroindole



TATTGGGCGCCAGGGTGGTTTTT
TATTGGGCGXCAMGGXGGTTTTT
X = 5-nitoindole or 3-nitropyrole
M = mutation site

FIGURE 74

	All Comp	1Base Mismatch	Allcomp 2NI	1Base Mismatch 2NI	Allcomp 1NP	1Base Mismatch 1NP
Invade #	67	114	115	116	112	113

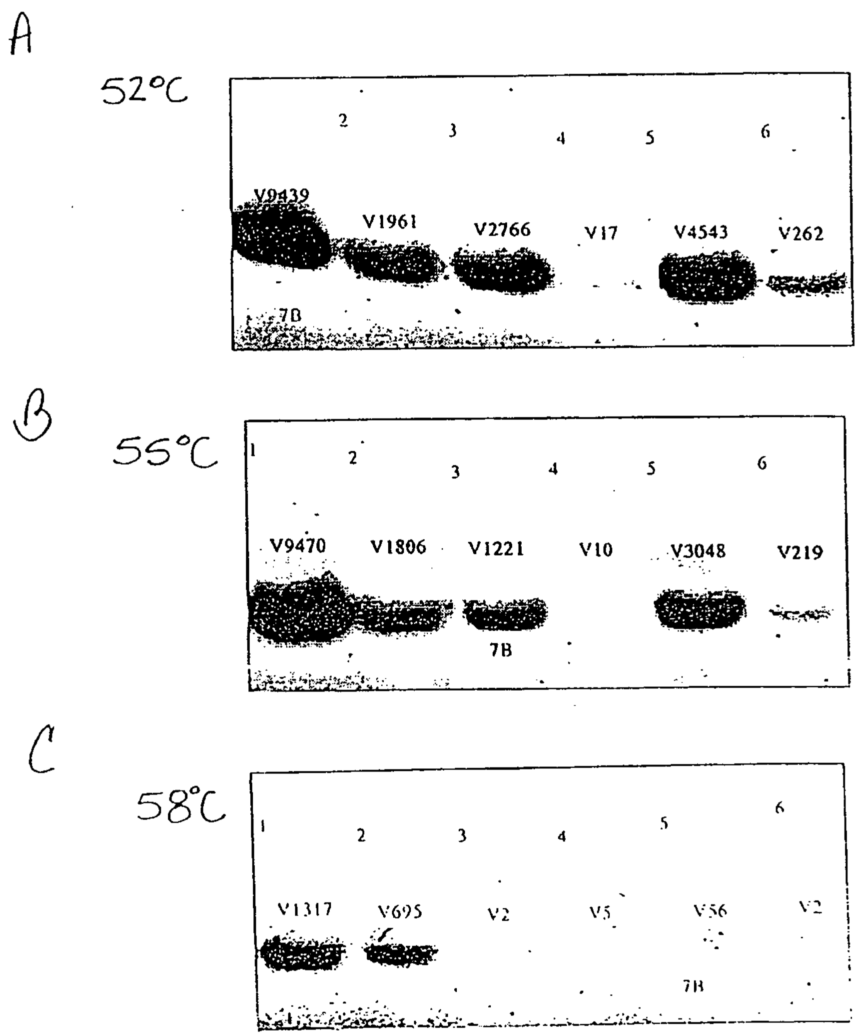


FIGURE 75

FIGURE 76

Codon 13
Point Mutation

166 (wt)
3'-CAGTTCCGTGAGAACGGATGCGGTGGTCGAGGTTGATGGTGTTCAAATATAAGTC-5'

165 (mutant)
3'-CAGTTCCGTGAGAACGGATGCAGTGGTCGAGGTTGATGGTGTTCAAATATAAGTC-5'

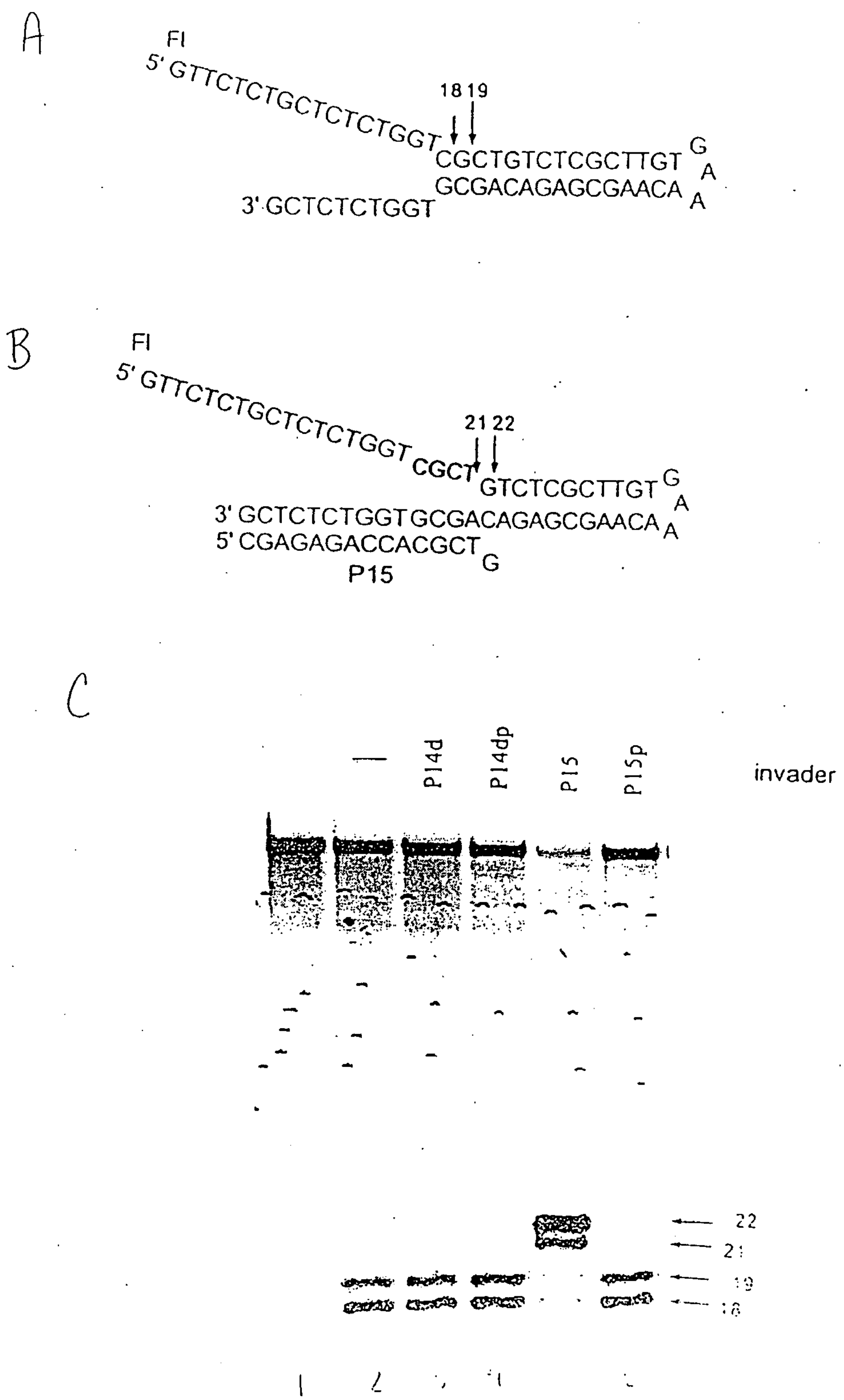
5'-CY3-T-T-CACCAG 161

3'-GCAGTTCCGTGAGAACGGATGCXGTGGTCGAGGTTGATGGTGTTCAAATATAAGTC-5'
162 ► <u>GCTCAAGGC</u>ACTCTTGCCTACGA **<u>CTCCAACTA</u>CCACAAGTTTATATTCAG ◄ 164**

FIGURE 77

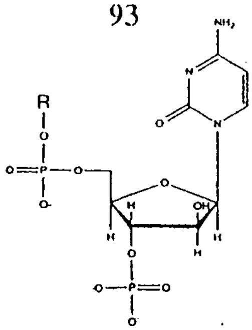
Temp →		47°		50°		53°		56°	
Target →	—	165	166	165	166	165	166	165	166

FIGURE 78


A
Fl
5' GTTCTCTGCTCTCTGGT
18 19
CGCTGTCTCGCTTGT G
A
GCGACAGAGCGAACA A
3' GCTCTCTGGT
B
Fl
5' GTTCTCTGCTCTCTGGT CGCT
21 22
GTCTCGCTTGT G
A
3' GCTCTCTGGT GCGACAGAGCGAACA A
5' CGAGAGACCACGCT G
P15
C
—
P14d
P14dp
P15
P15p
invader
22
21
19
18

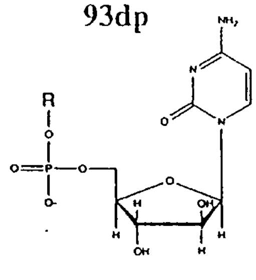
Invader 3'-end Substituents

93



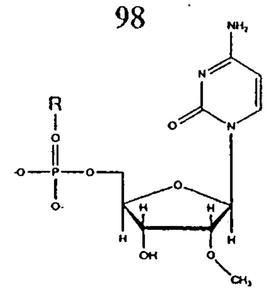
3'Phosphate C Arabinose

93dp



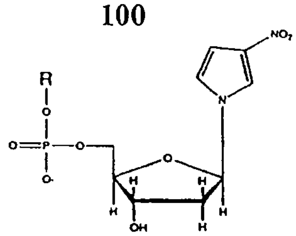
3'-Cytosine Arabinose

98



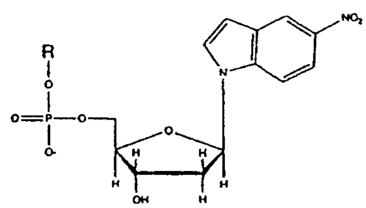
2'-Omethyl Cytosine

100



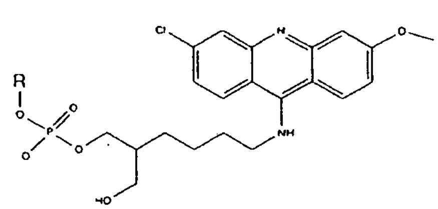
3-nitropyrrole

101



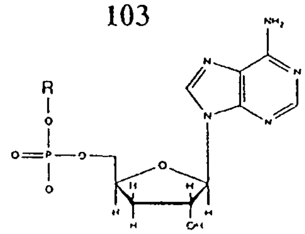
5-nitroindole

102



Acridine

103



3'-deoxy-adenine

FIGURE 79

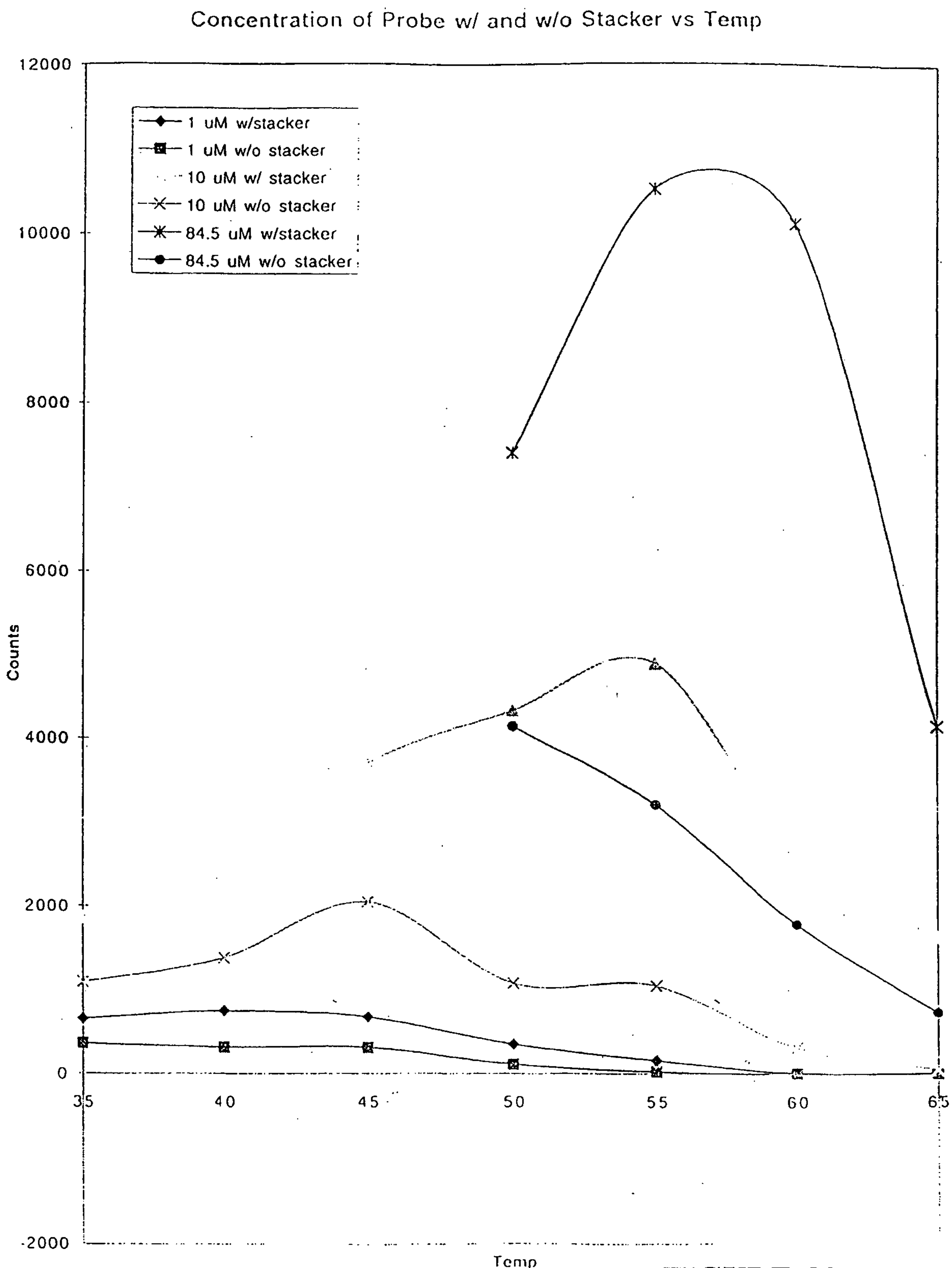

Concentration of Probe w/ and w/o Stacker vs Temp
12000
10000
8000
6000
4000
2000
0
-2000
Counts
35
40
45
50
55
60
65
Temp
1 uM w/stacker
1 uM w/o stacker
10 uM w/ stacker
10 uM w/o stacker
84.5 uM w/stacker
84.5 uM w/o stacker

FIGURE 80

FIGURE 81

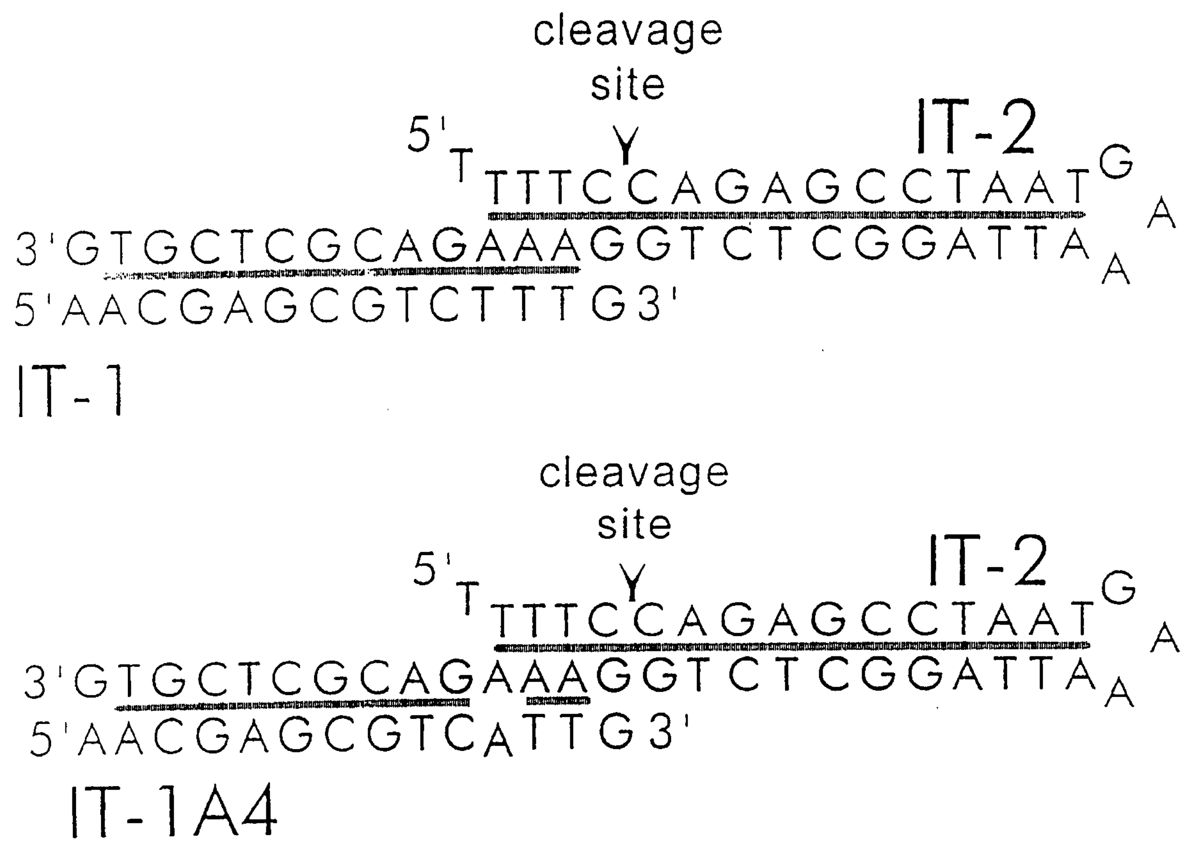

cleavage
site
5'T TTTCCAGAGCCTAAT G A
IT-2
3'GTGCTCGCAGAAAGGTCTCGGATTA A
5'AACGAGCGTCTTTG3'
IT-1
cleavage
site
5'T TTTCCAGAGCCTAAT G A
IT-2
3'GTGCTCGCAGAAAGGTCTCGGATTA A
5'AACGAGCGTCATTG3'
IT-1A4

FIGURE 82

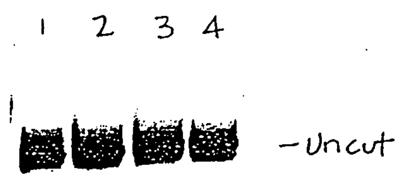


FIGURE 83

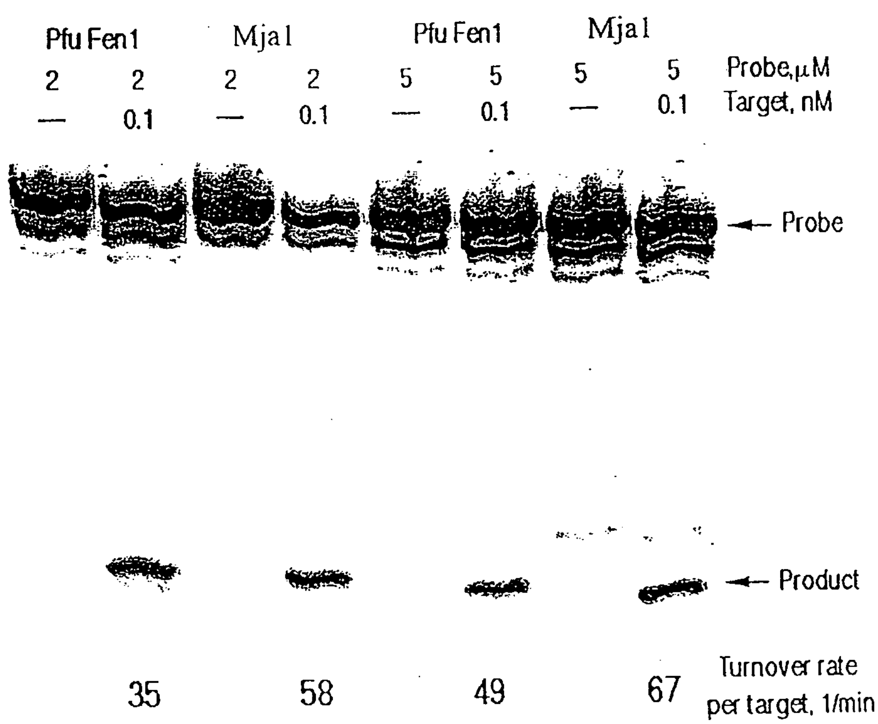

Pfu Fen1
Mja1
Pfu Fen1
Mja1
2 2 2 2 5 5 5 5 Probe,μM
— 0.1 — 0.1 — 0.1 — 0.1 Target, nM
Probe
Product
35 58 49 67
Turnover rate per target, 1/min

FIGURE 84

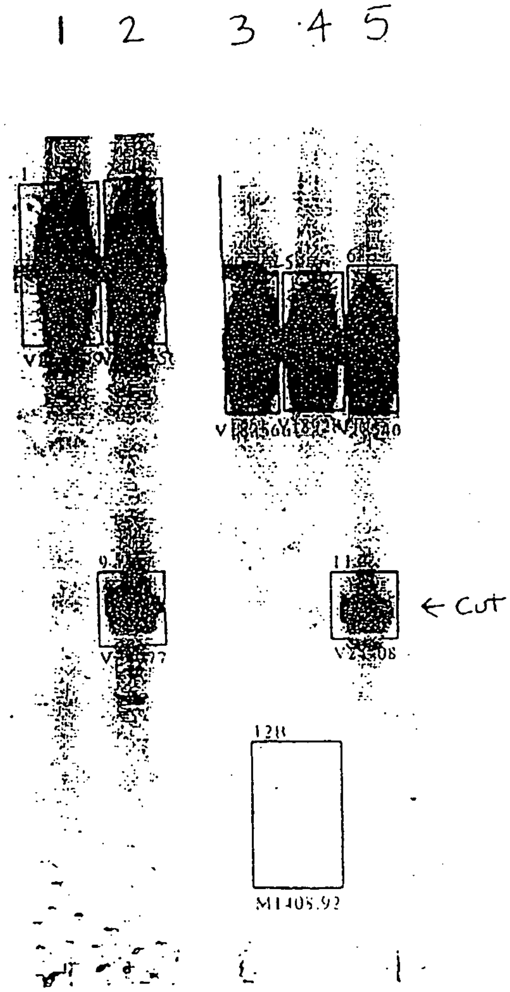
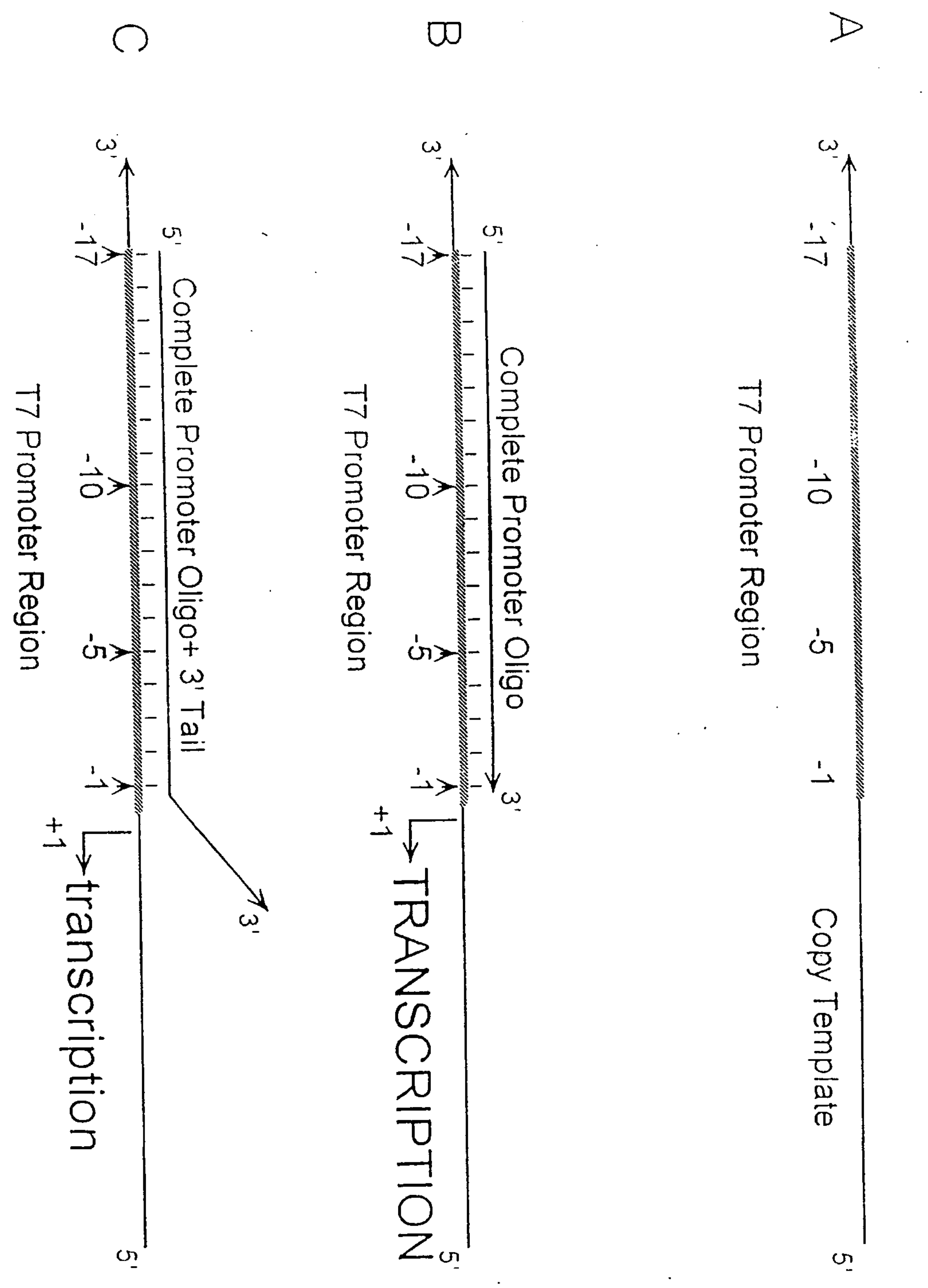


FIGURE 85


A
3'
-17
-10
-5
-1
5'
T7 Promoter Region
Copy Template
B
5'
Complete Promoter Oligo
3'
-17
-10
-5
-1
+1
TRANSCRIPTION
T7 Promoter Region
C
Complete Promoter Oligo+ 3' Tail
transcription

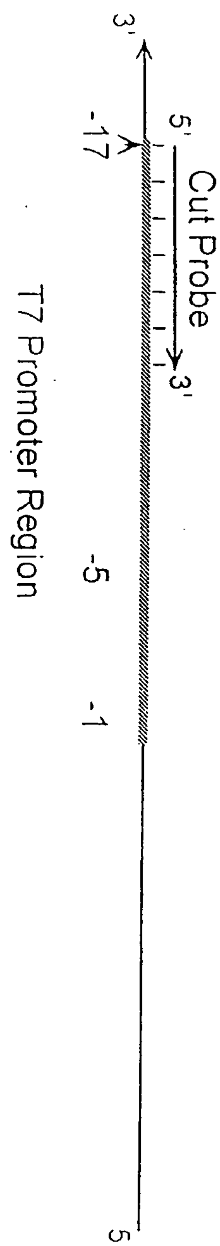


FIGURE 86A

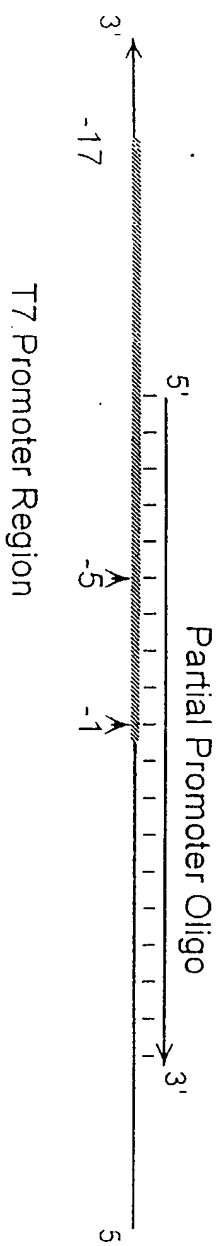


FIGURE 86B

FIGURE 86C

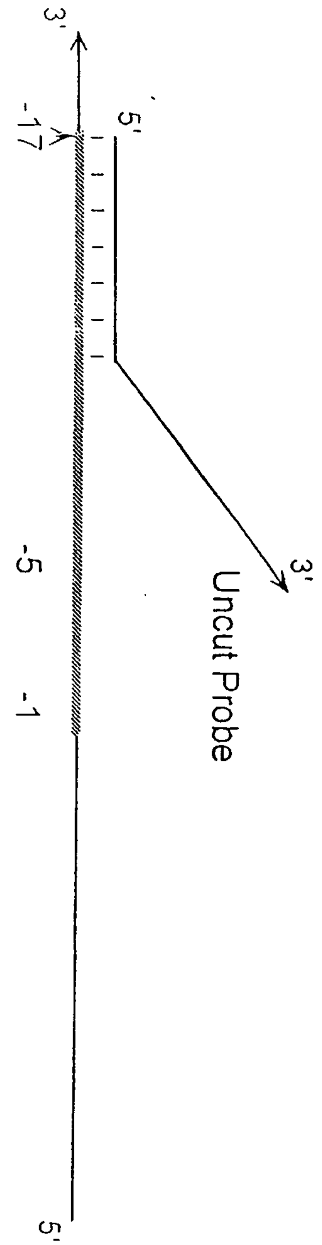


FIGURE 86D

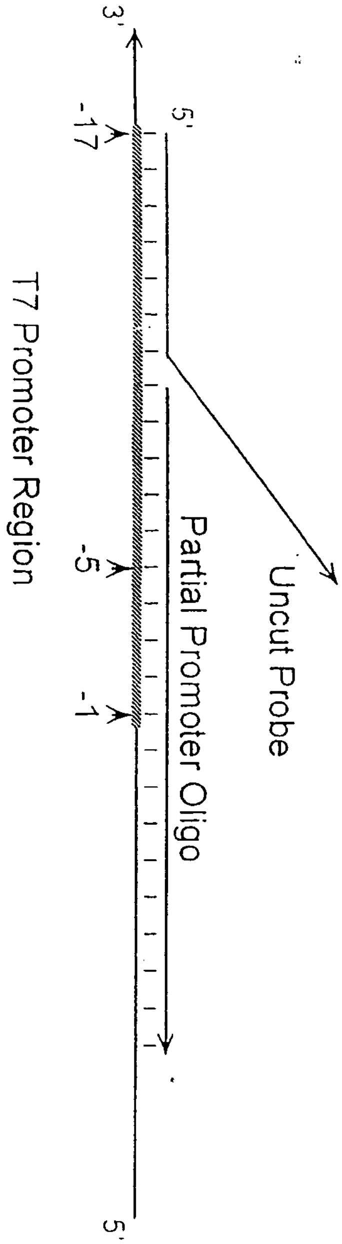


FIGURE 87

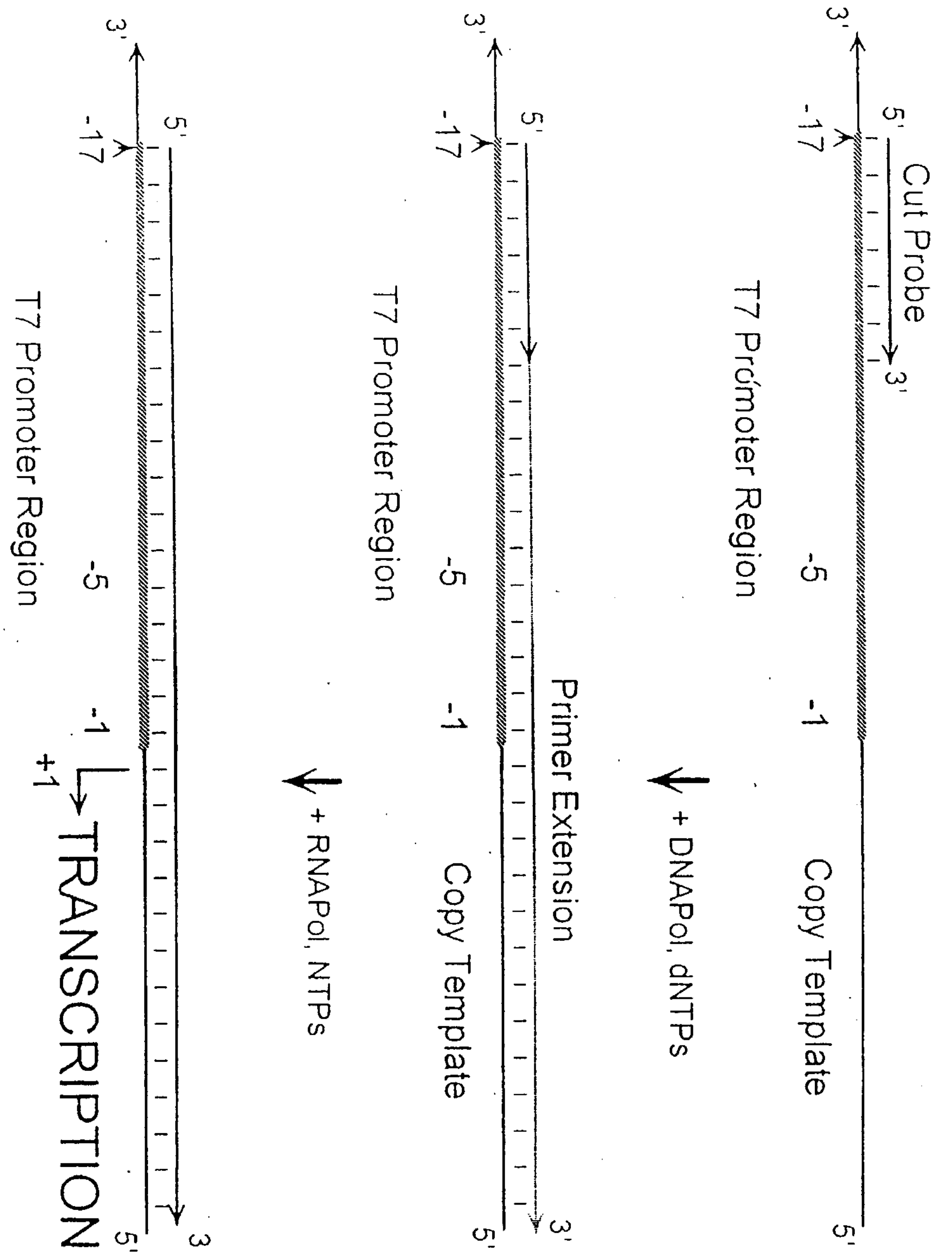

3'
5'
Cut Probe
3'
-17
-5
-1
T7 Prómoter Region
Copy Template
5'
+ DNAPol, dNTPs
3'
5'
-17
-5
-1
Primer Extension
T7 Promoter Region
Copy Template
3'
5'
+ RNAPol, NTPs
3'
5'
-17
-5
-1
+1
T7 Promoter Region
TRANSCRIPTION
3'
5'

FIGURE 88A

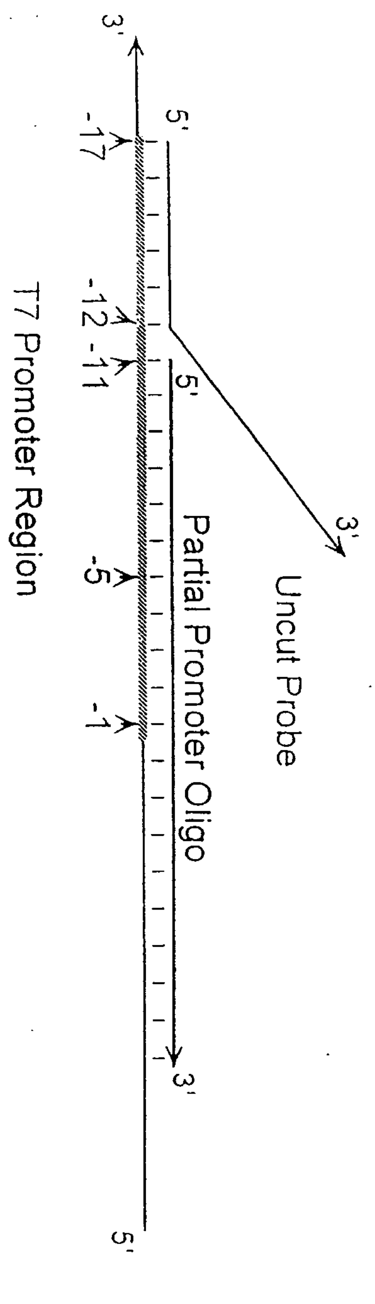


FIGURE 88B

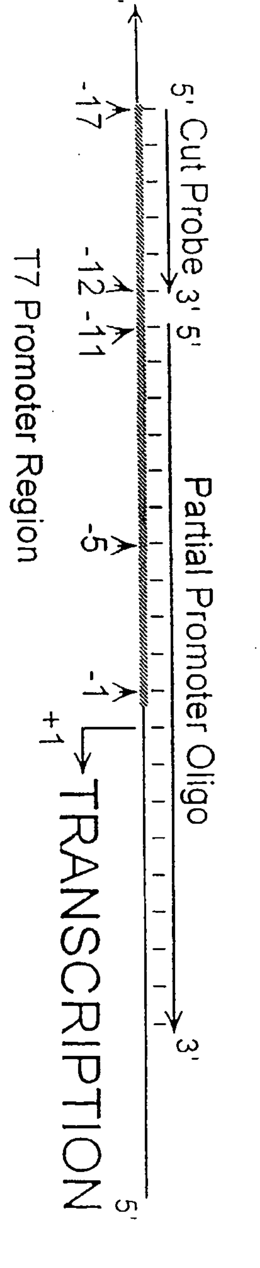


FIGURE 89

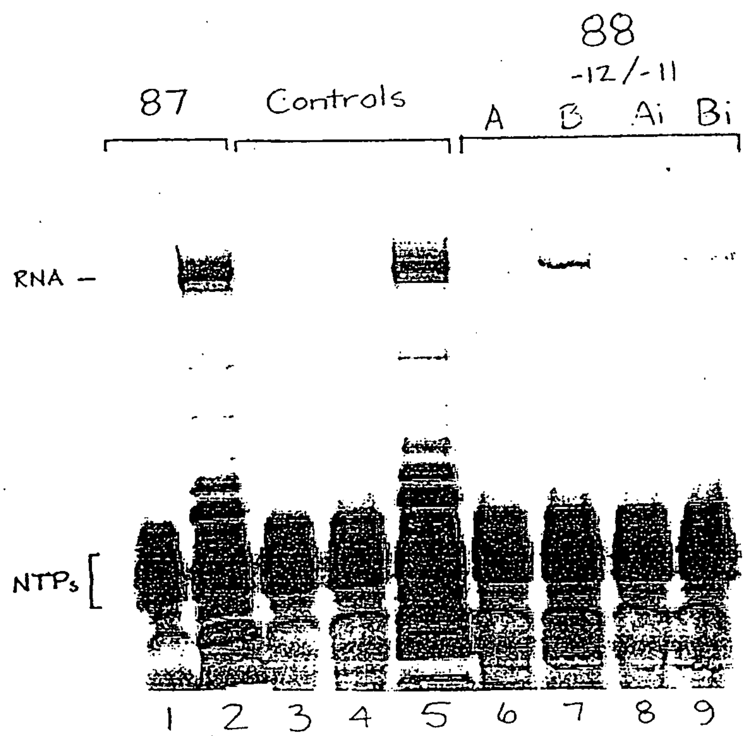
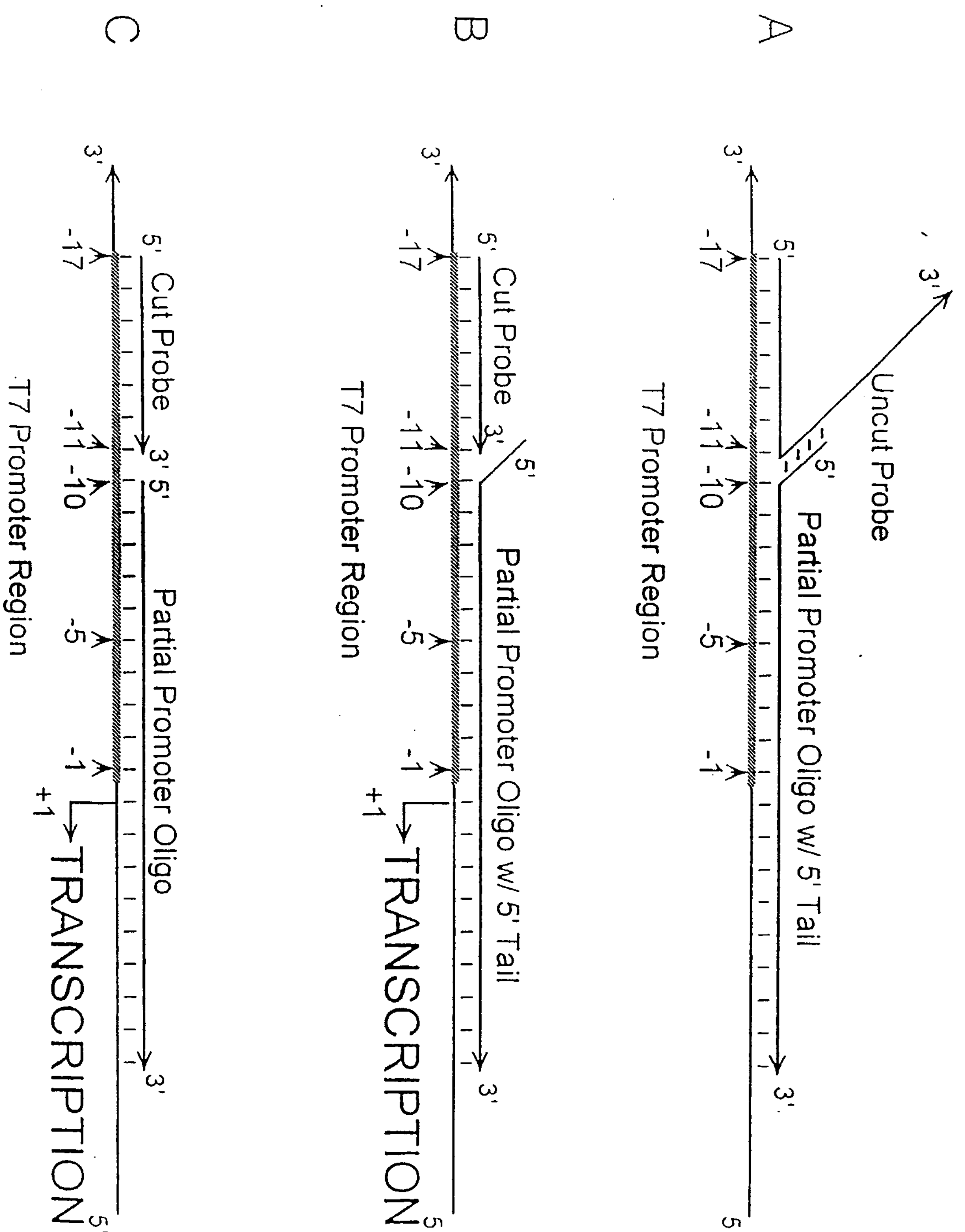


FIGURE 90


A
Uncut Probe
Partial Promoter Oligo w/ 5' Tail
T7 Promoter Region
-17
-11 -10
-5
-1
B
Cut Probe
Partial Promoter Oligo w/ 5' Tail
T7 Promoter Region
+1
TRANSCRIPTION
C
Cut Probe
Partial Promoter Oligo
T7 Promoter Region
+1
TRANSCRIPTION

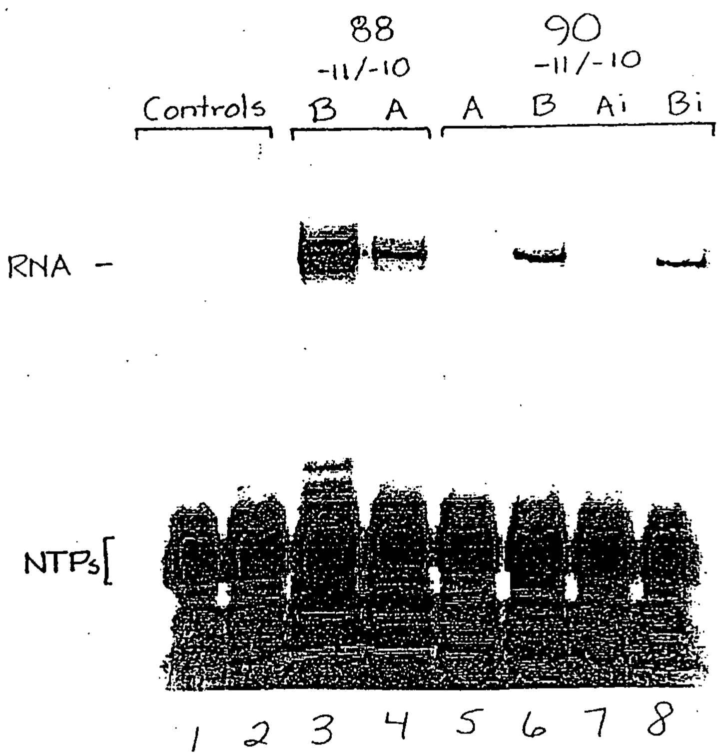


FIGURE 91

FIGURE 92

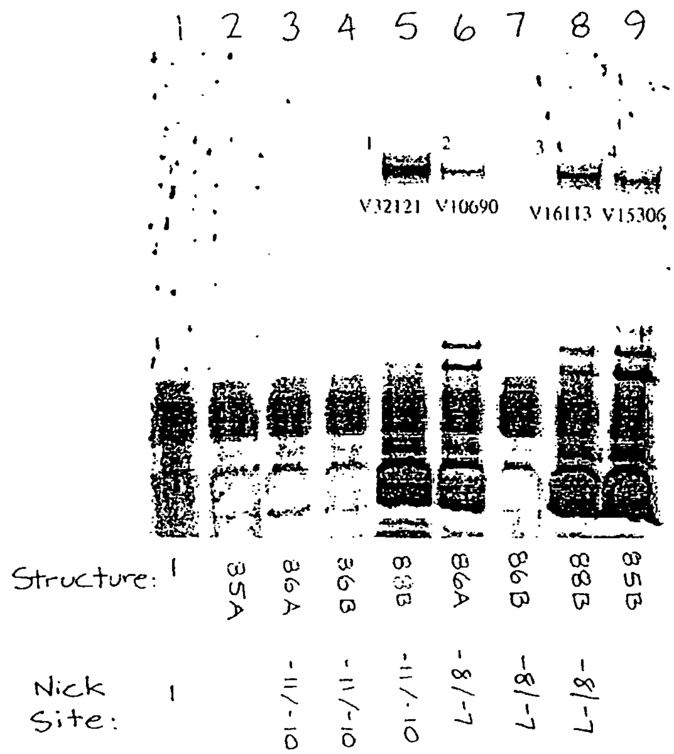
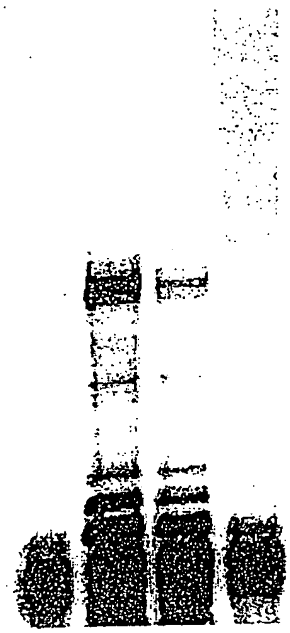

1 2 3 4 5 6 7 8 9
1 2 3 4
V32121 V10690 V16113 V15306
Structure: 1 85A 86A 86B 83B 86A 86B 88B 85B
Nick Site: 1 -11/-10 -11/-10 -11/-10 -8/-7 -8/-7 -8/-7

FIGURE 93

1 2 3 4

RNA —



1 85A

3'-gctttaattatgctgagtgatatcccagaagatacctccagttttgt-5'

-23 -20 -15 -10 -5 -1

T7 promoter region

2 85B

5'-cgaaattaatacgactcactata-3'

3'-gctttaattatgctgagtgatatcccagaagatacctccagttttgt-5'

-23 -20 -15 -10 -5 -1

T7 promoter region

3 85C

#073-065

5'-cgaaattaatacgactcactata cccagaa -3'

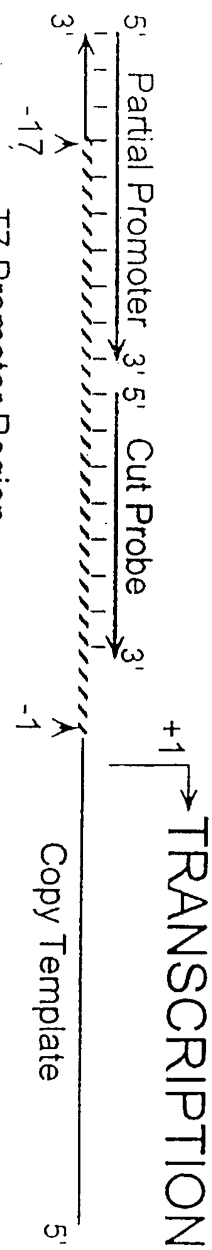
3'-gctttaattatgctgagtgatatcccagaagatacctccagttttgt-5'

-23 -20 -15 -10 -5 -1

T7 promoter region

4 No DNA

FIGURE 94


5'
Partial Promoter
3' 5'
Cut Probe
3'
3'
-17
-1
T7 Promoter Region
+1
TRANSCRIPTION
Copy Template
5'

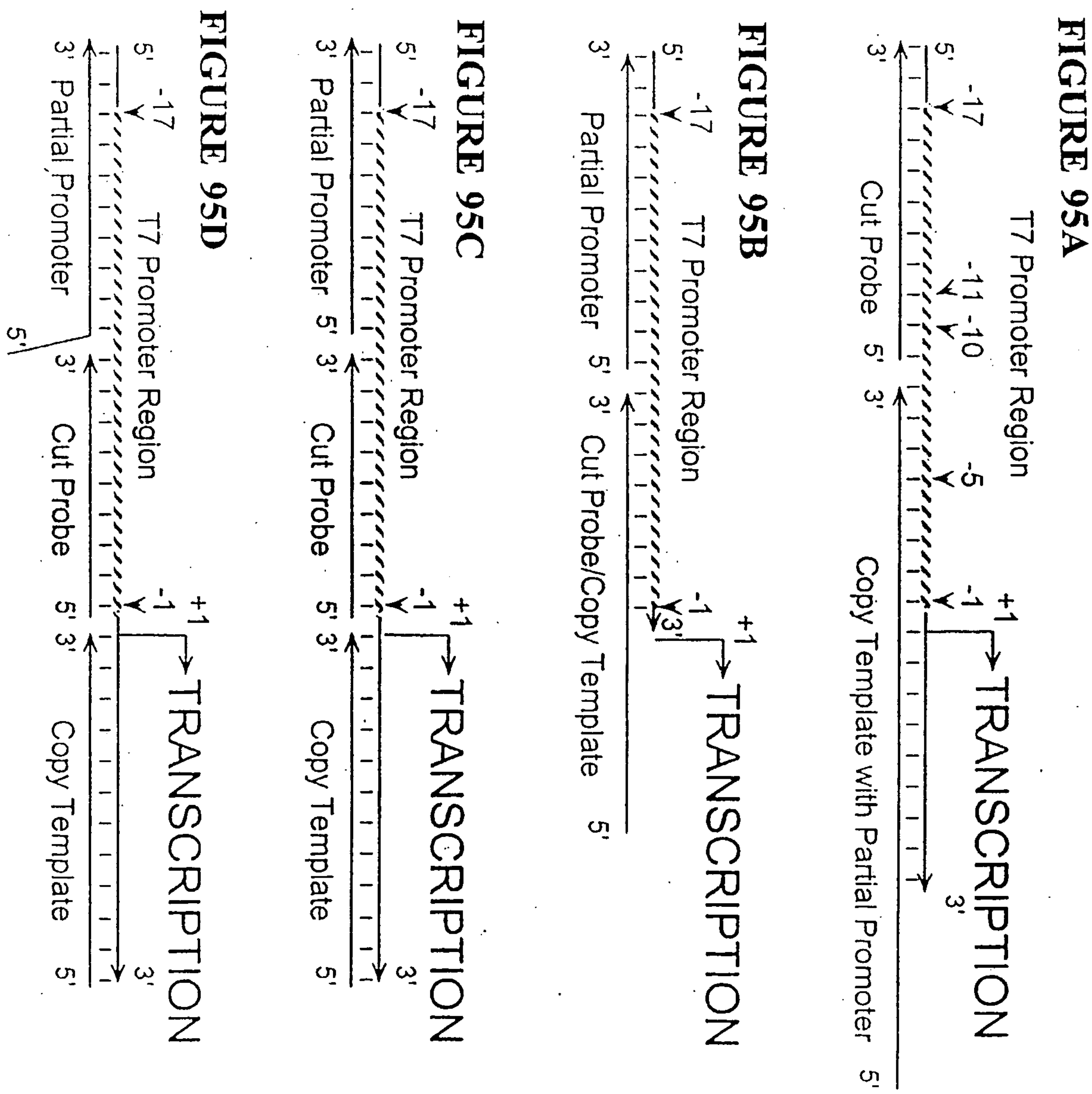


FIGURE 95A

FIGURE 95B

FIGURE 95C

FIGURE 95D

FIGURE 96

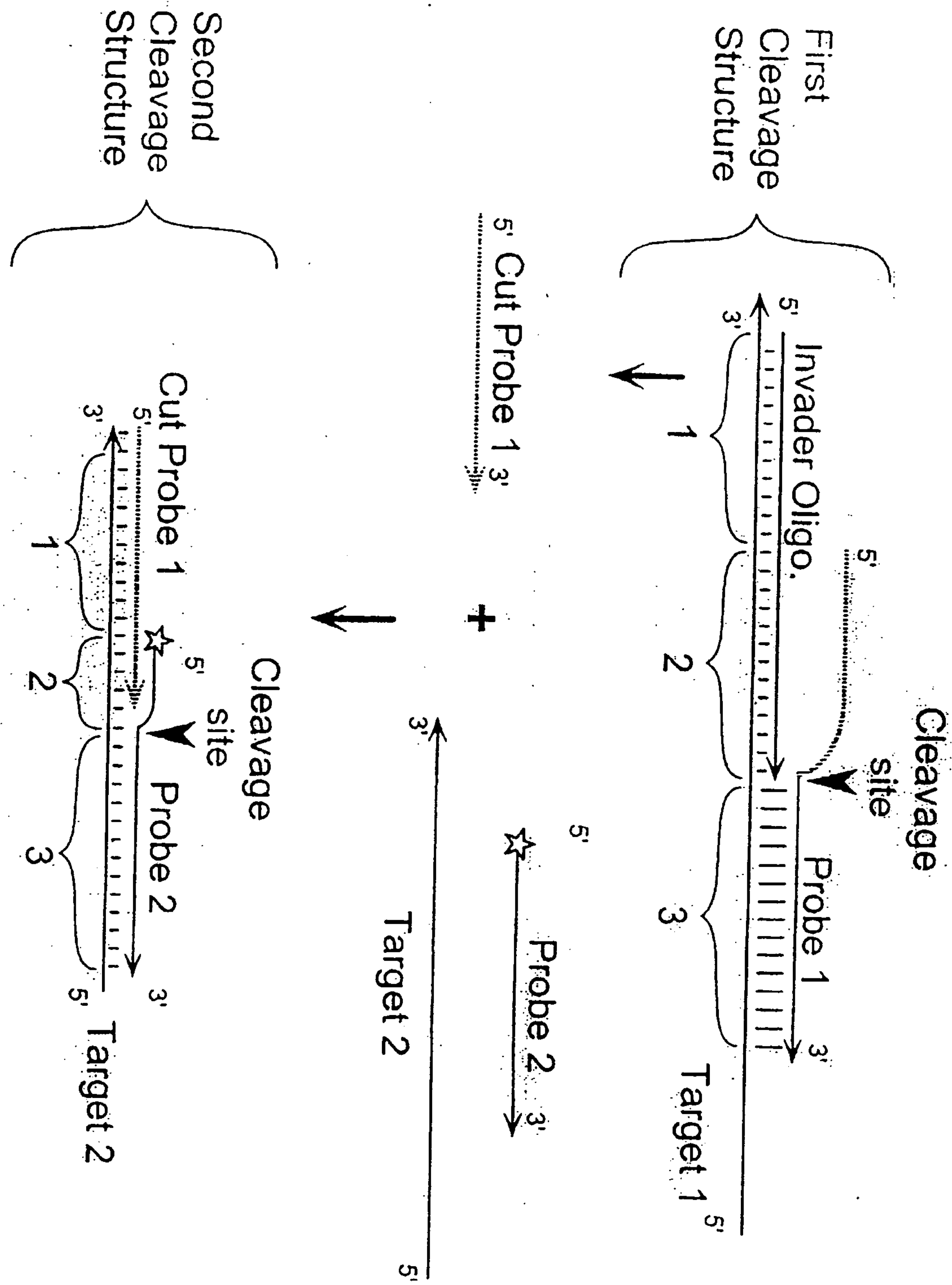

First Cleavage Structure
Invader Oligo.
Cleavage site
Probe 1
Target 1
1
2
3
5' Cut Probe 1 3'
+
Probe 2
Target 2
Second Cleavage Structure
Cut Probe 1
Cleavage site
Probe 2
Target 2
1
2
3

FIGURE 97

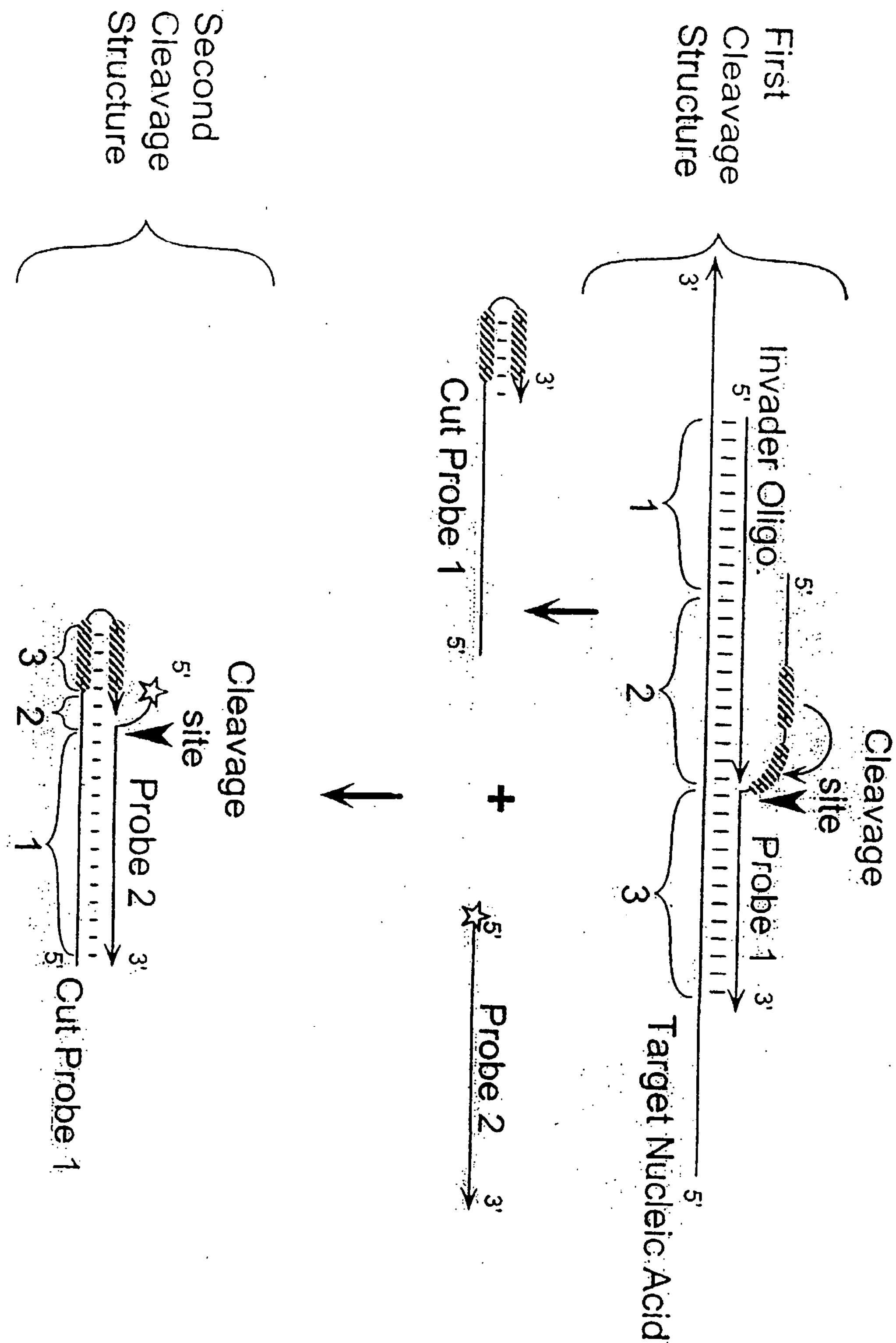

First Cleavage Structure
Invader Oligo.
Cleavage site
Probe 1
Target Nucleic Acid
1
2
3
3'
5'
Cut Probe 1
Probe 2
Second Cleavage Structure
Cleavage site
Probe 2
Cut Probe 1

FIGURE 98

PR1 probe

```
         Cleavage site
              |
5'FITTTTCCAGAGCCTAAT G3'
```

IT3 Invader-Target

```
 A ACGAGCGTCTTT G3'
A
 G TGCTCGCAGAAAGGTCTCGGATTAATTTTTTTTT5'
```

IT3-8 Invader-Target

```
 A AGCGTCTT G3'
A
 G TCGCAGAAGGTCTCGGATTAATTTTTTTTT5'
```

IT3-6 Invader-Target

```
 A CGTCTT G3'
A
 G GCAGAAGGTCTCGGATTAATTTTTTTTT5'
```

IT3-4 Invader-Target

```
 A TCTT G3'
A
 G AGAAGGTCTCGGATTAATTTTTTTTT5'
```

IT3-3 Invader-Target

```
 A CTT G3'
A
 G GAAGGTCTCGGATTAATTTTTTTTT5'
```

IT3-0 Invader-Target

```
3'GAAGGTCTCGGATTAATTTTTTTTT5'
```

FIGURE 99

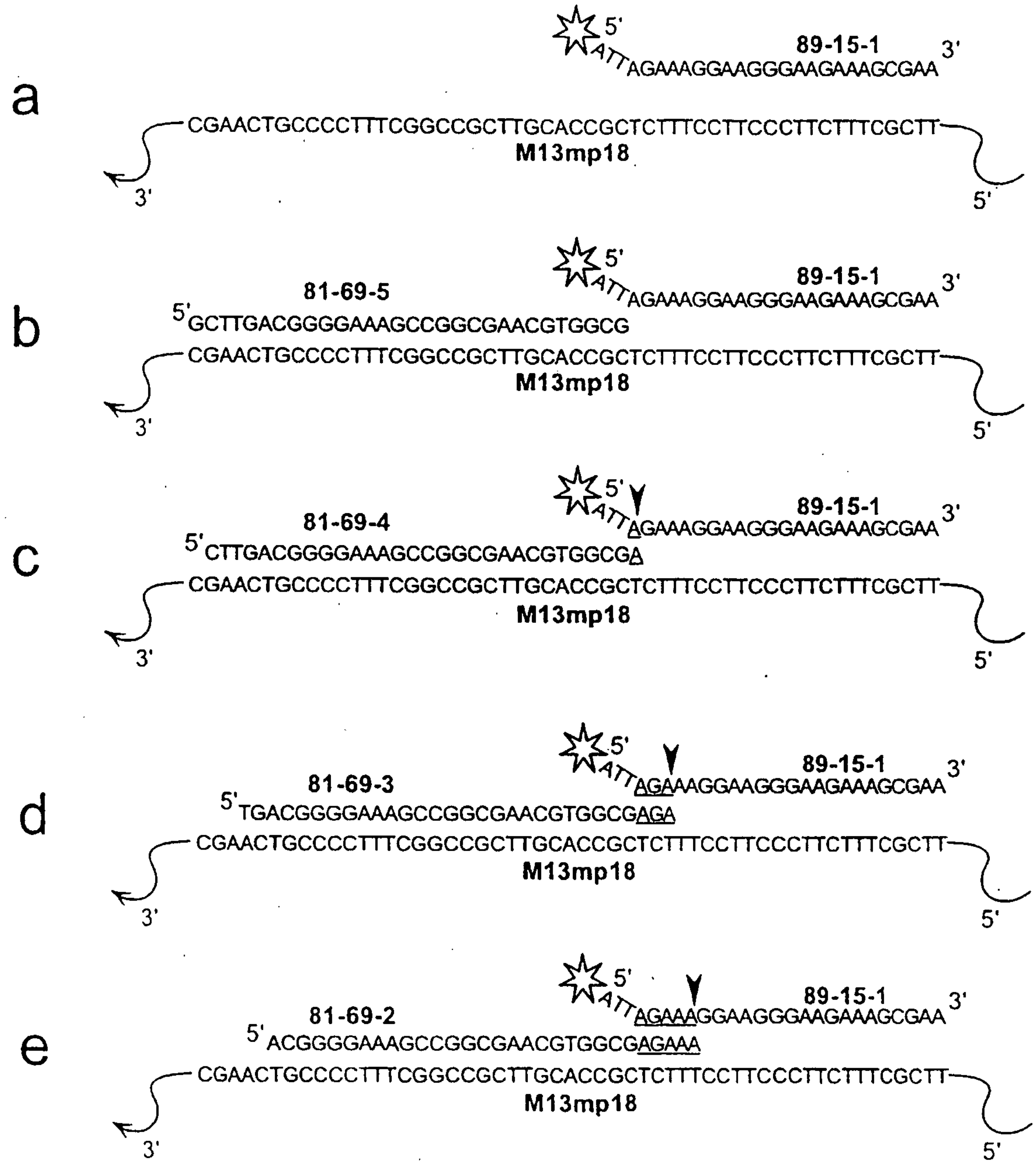

a
5'
89-15-1
3'
ATTAGAAAGGAAGGGAAGAAAGCGAA
CGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTT
M13mp18
3'
5'
b
81-69-5
5' GCTTGACGGGGAAAGCCGGCGAACGTGGCG
89-15-1
ATTAGAAAGGAAGGGAAGAAAGCGAA
CGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTT
M13mp18
c
81-69-4
5' CTTGACGGGGAAAGCCGGCGAACGTGGCGA
89-15-1
ATTAGAAAGGAAGGGAAGAAAGCGAA
CGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTT
M13mp18
d
81-69-3
5' TGACGGGGAAAGCCGGCGAACGTGGCGAGA
89-15-1
ATTAGAAAGGAAGGGAAGAAAGCGAA
CGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTT
M13mp18
e
81-69-2
5' ACGGGGAAAGCCGGCGAACGTGGCGAGAAA
89-15-1
ATTAGAAAGGAAGGGAAGAAAGCGAA
CGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCCTTCCCTTCTTTCGCTT
M13mp18

FIGURE 100

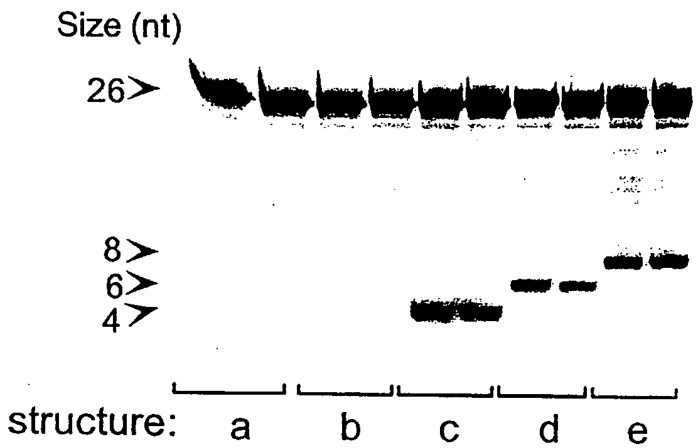
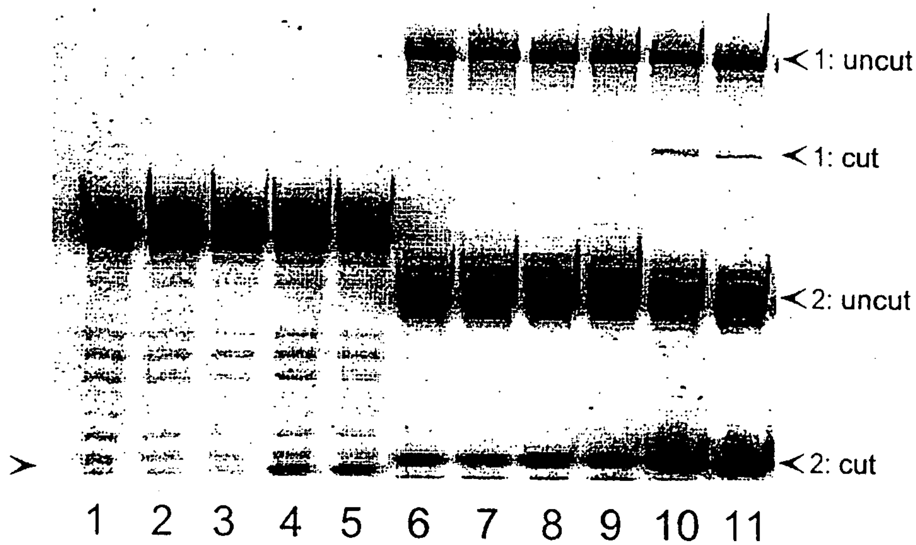
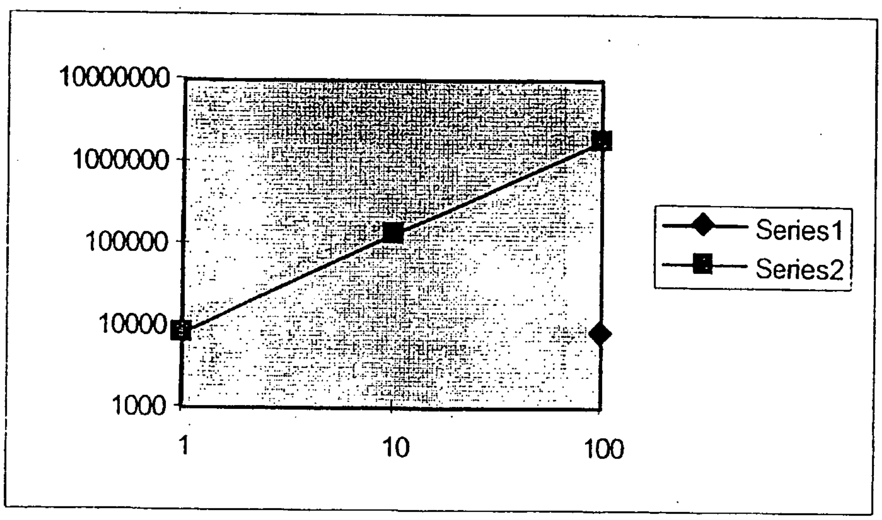


FIGURE 101

a



b



attomoles/target	basic invader	invader sqrd
1		8386
10		133185
100	8512	1862211

FIGURE 102

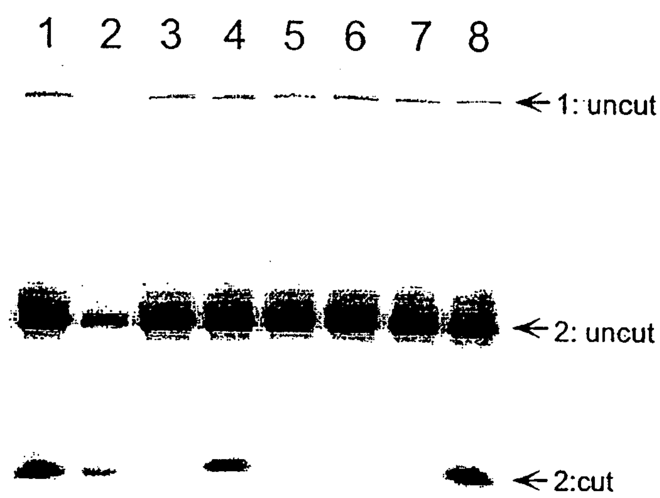
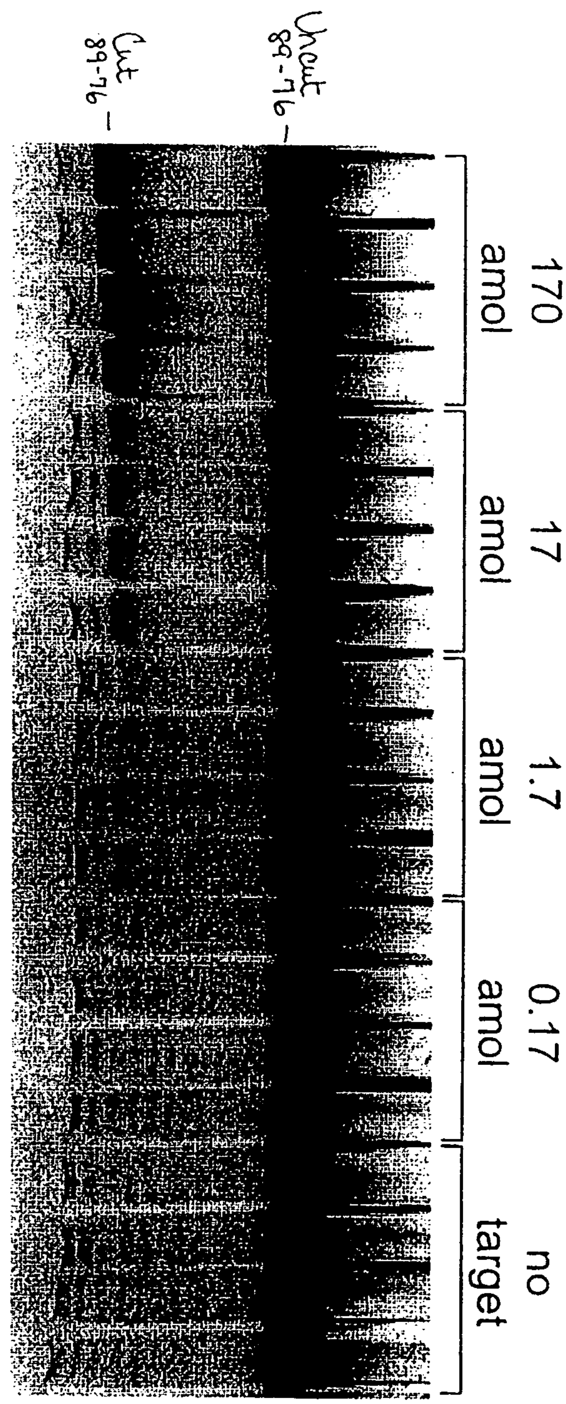


FIGURE 103

```
                                       Cleavage
                                         site
                                    5'Fl-TTT ▼        89-76
     89-44                                   CCCTCCTCCTCTTCC-3'
   5'ACACAGTGTCCTCCCGCTCCTCCTGAGCAA
3'-ACTGTCTGTGTCACAGGAGGGCGAGGAGGACTCGTGGGAGGAGGAGAAGGAGTA-5'
   |                                                     |
   3110                                                  3057
```

HCMV Target Sequence

FIGURE 104


Uncut 89-76
Cut 89-76
170 amol
17 amol
1.7 amol
0.17 amol
no target

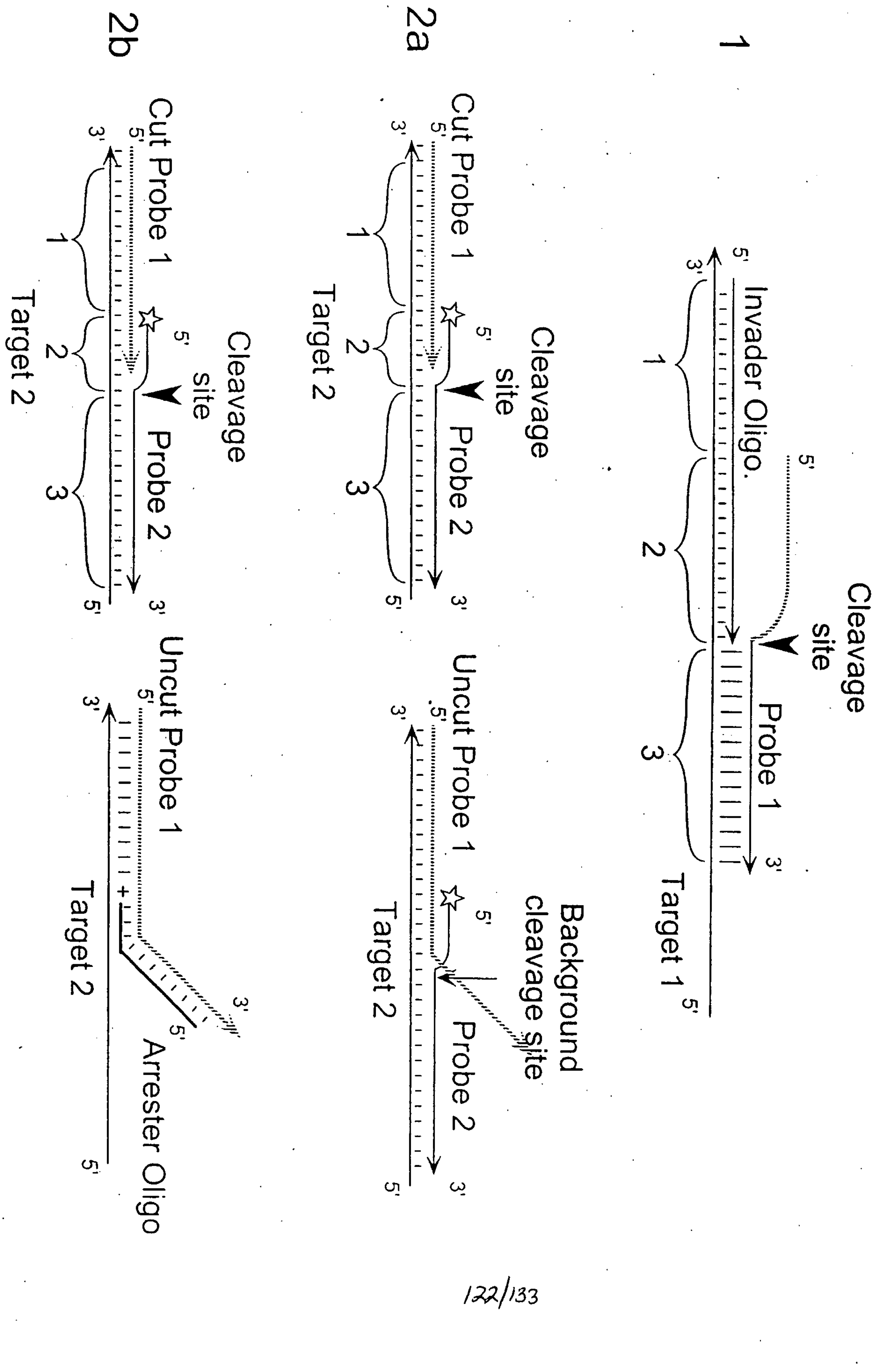
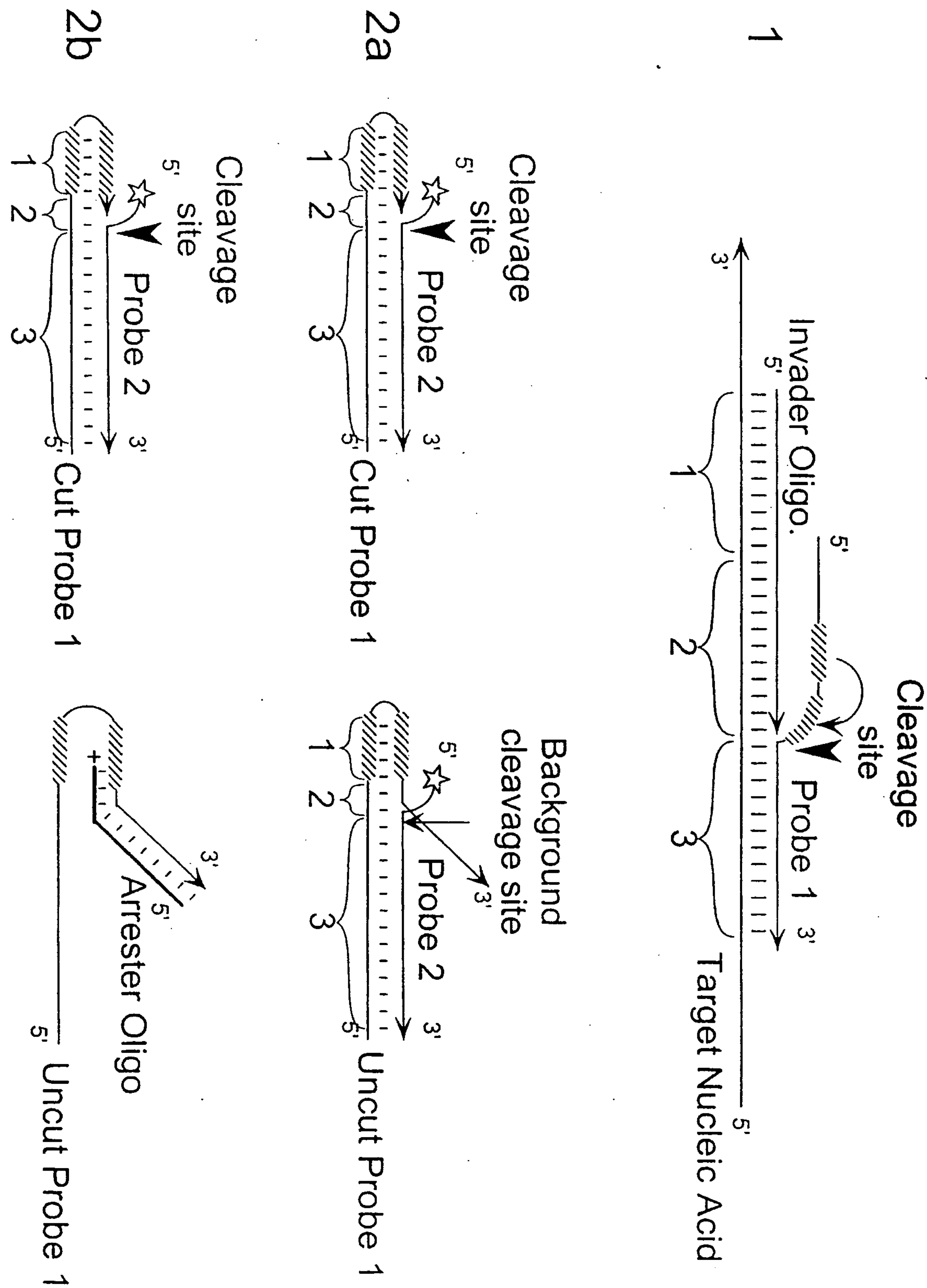
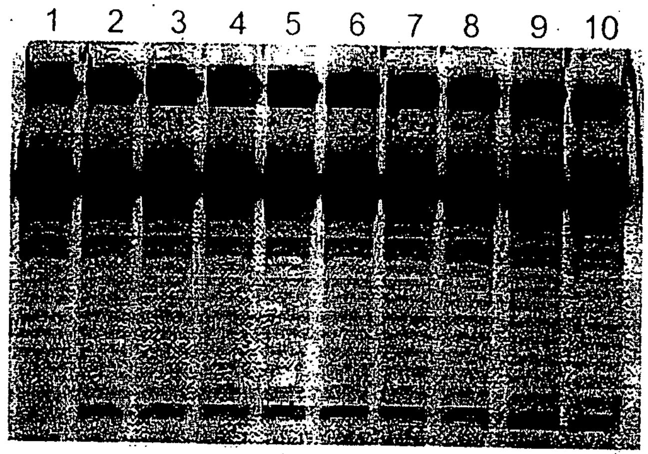

1
Invader Oligo.
Probe 1
Cleavage site
Target 1
2a
Cut Probe 1
Probe 2
Cleavage site
Target 2
Uncut Probe 1
Background cleavage site
Probe 2
Target 2
2b
Cut Probe 1
Probe 2
Cleavage site
Target 2
Uncut Probe 1
Arrester Oligo
Target 2

FIGURE 105

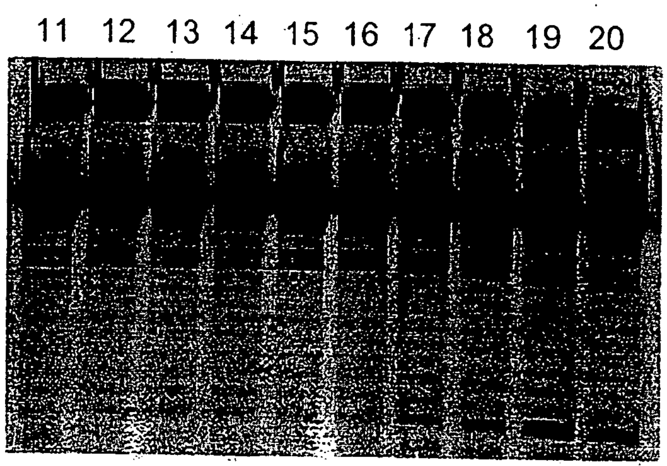
FIGURE 106


1
Cleavage site
Invader Oligo.
Probe 1
Target Nucleic Acid
5'
3'
1
2
3
2a
Cleavage site
Probe 2
Cut Probe 1
Background cleavage site
Uncut Probe 1
2b
Arrester Oligo
+

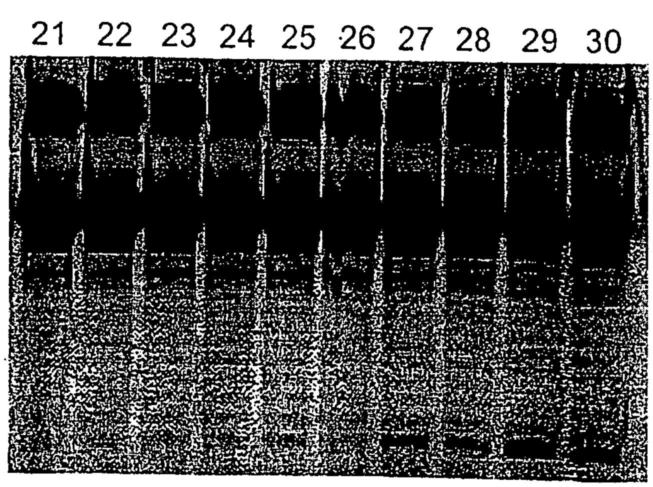
A


1 2 3 4 5 6 7 8 9 10

B


11 12 13 14 15 16 17 18 19 20

C


21 22 23 24 25 26 27 28 29 30

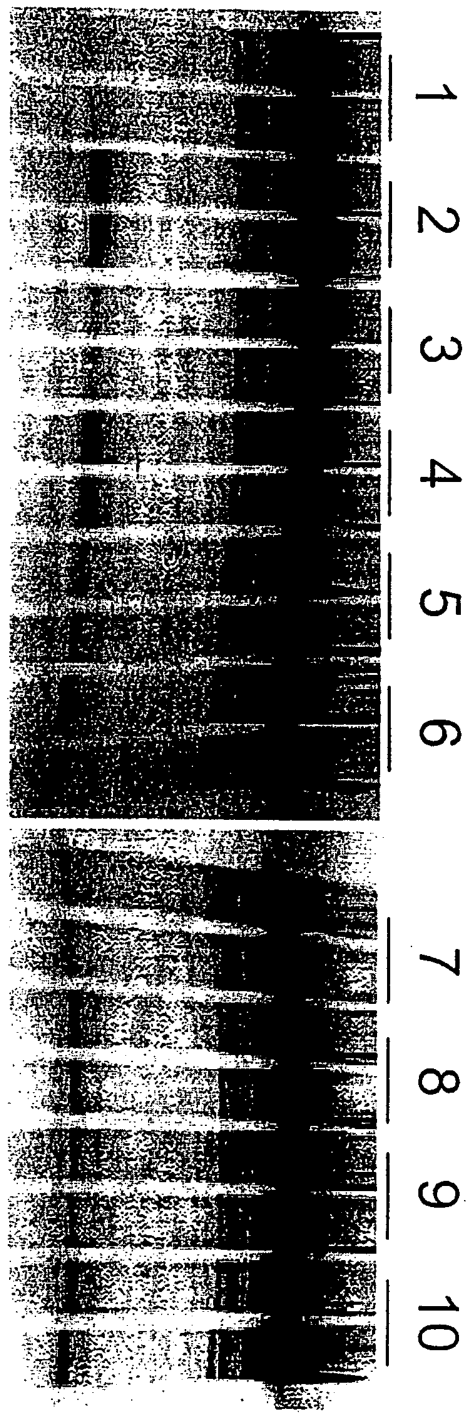


FIGURE 108A

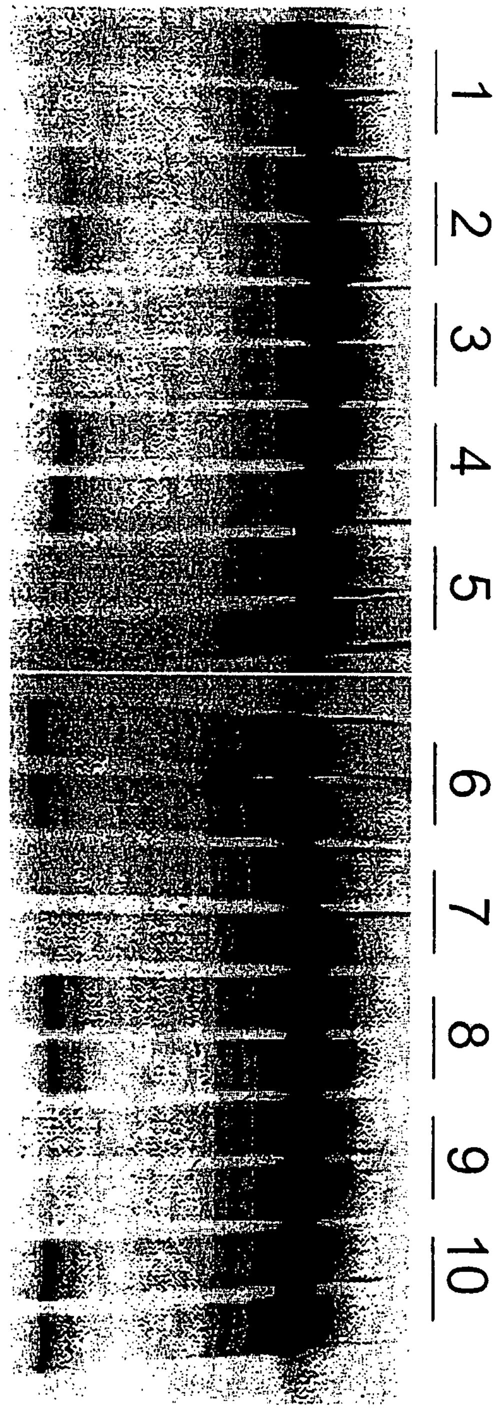


FIGURE 108B

FIGURE 108C

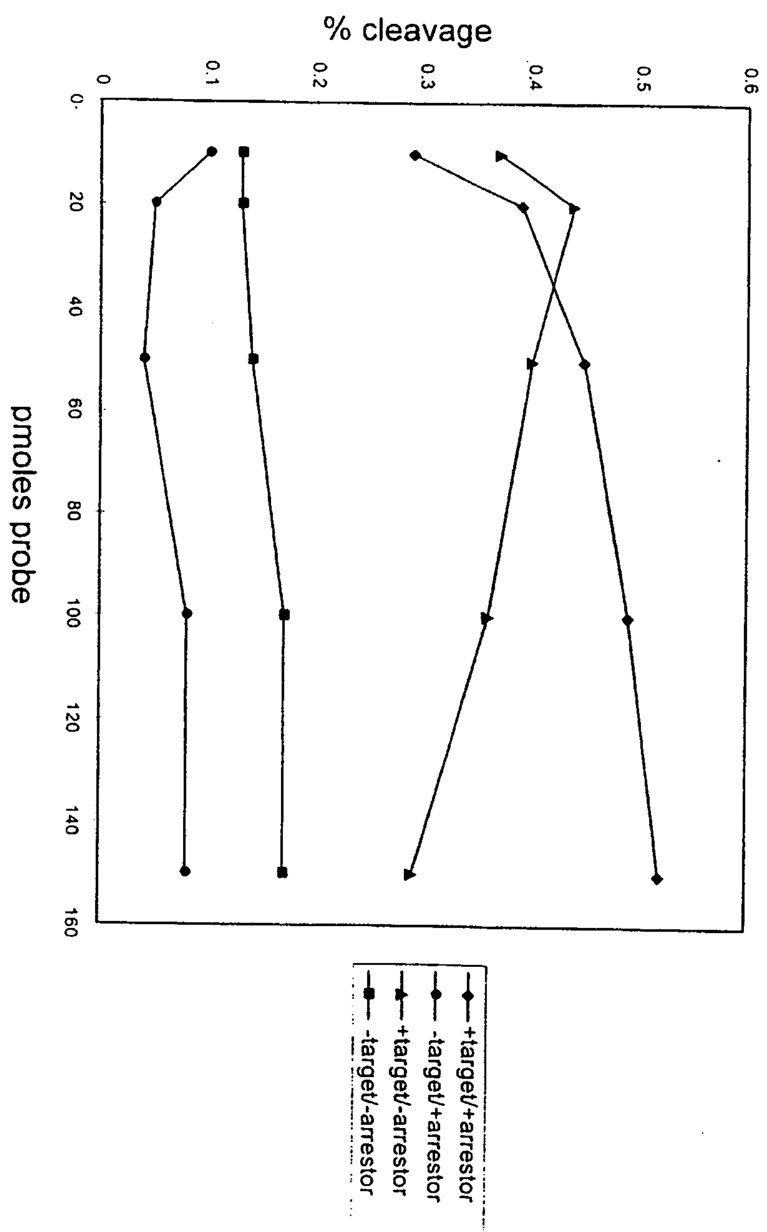
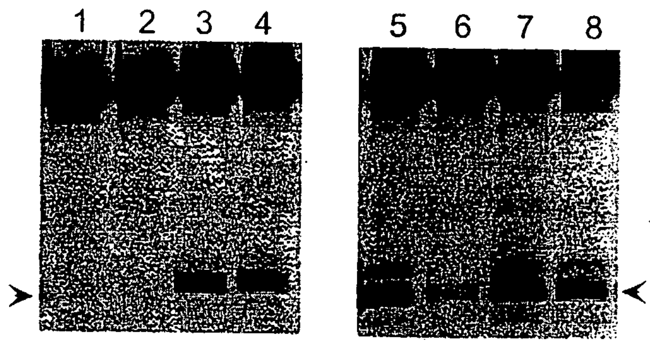

% cleavage
0
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
120
140
160
pmoles probe
+target/+arrestor
-target/+arrestor
+target/-arrestor
-target/-arrestor

FIGURE 109A



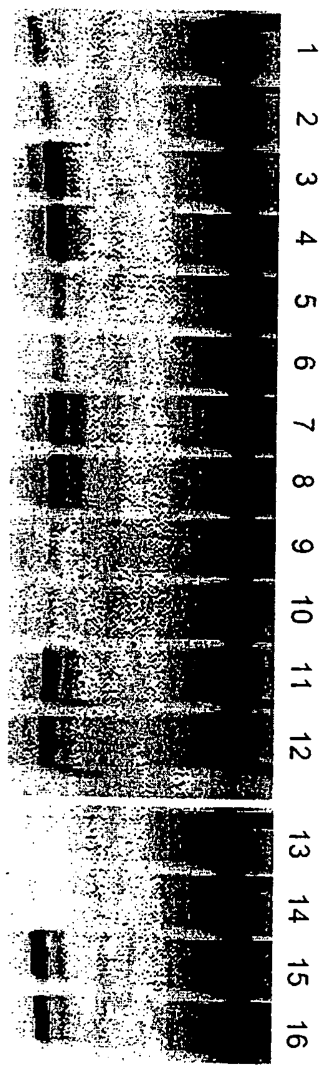


FIGURE 109B

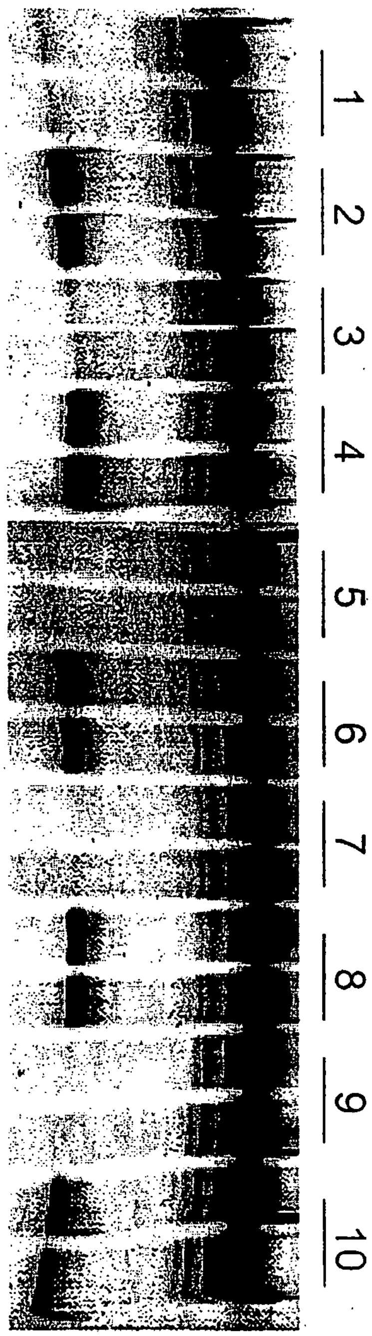


FIGURE 110A

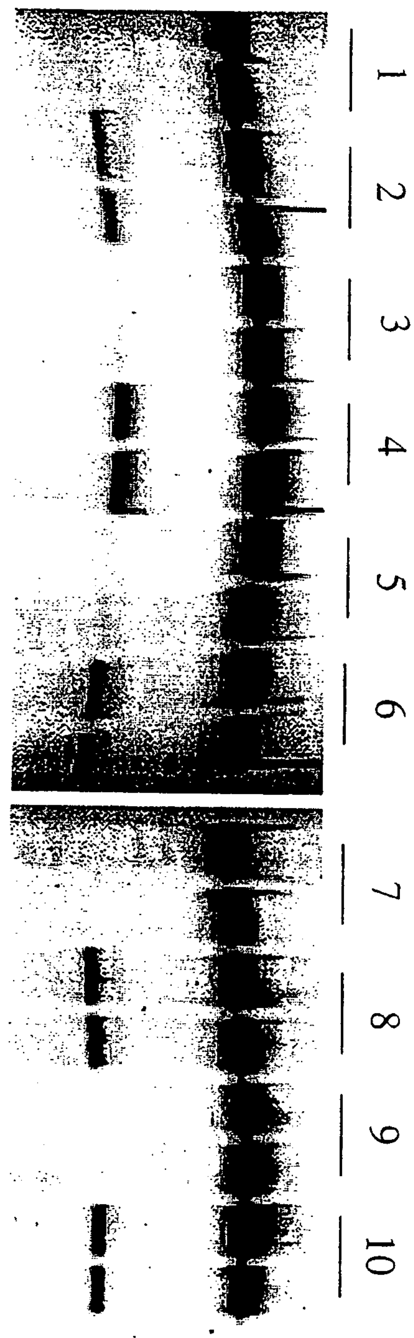


FIGURE 110B

FIGURE 110C

Primary Probe 241-95-02

5' AACGAGGCGCACCCACCCAAGGCACAGC-NH3+ 3'

Arrestors

3' NH3+GGGTGGGTTCCGTGTCG 5' 241-95-03

3' NH3+TGGGGTGGGTTCCGTGTCG 5' 241-95-04

3' NH3+TGCGGGGTGGGTTCCGTGTCG 5' 241-95-05

3' NH3+TGCGCGGGTGGGTTCCGTGTCG 5' 241-95-06

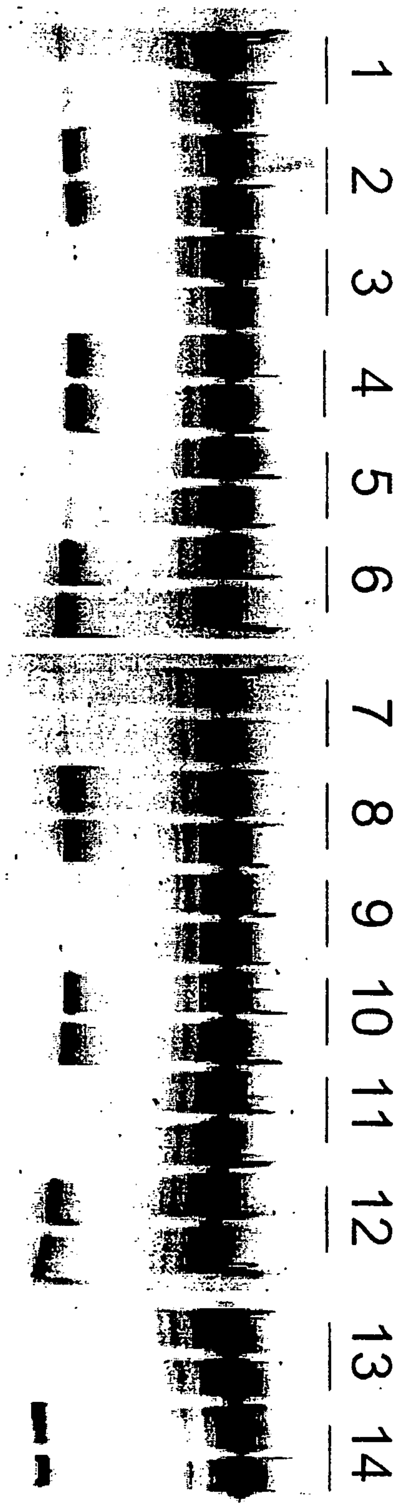


FIGURE 111

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/05809

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12Q 1/68; C12P 19/32
US CL : 435/6, 91.53
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.2, 91.51, 91.53, 195, 199; 536/24.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Searched inventors and keywords: nucleic acid hybridization and invader or cleave or helper or sequential in APS, CAPLUS, MEDLINE, SCISEARCH and WPIDS.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	LYAMICHEV et al.. Structure-Specific Endonucleolytic Cleavage of Nucleic Acids by Eubacterial DNA Polymerases. Science. 07 May 1993, Vol. 260, pages 778-783, see entire document.	1-68
Y	US 5,422,253 A (DAHLBERG et al.) 06 June 1995, see entire document.	1-68
Y	US 5,541,311 A (DAHLBERG et al.) 30 July 1996, see entire document.	1-68
Y	US 5,487,972 A (GELFAND et al.) 30 January 1996, see abstract, column 1, 5, 10, 12 and 25.	6-12,26-28,40,41-46,63,65-68
Y	US 5,601,976 A (YAMANE et al.) 11 February 1997, see abstract, column 1-2 and Figure 1.	13-19, 47-50

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 MAY 1998	09 JUL 1998

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	DEBRA SHOEMAKER
Facsimile No. (703) 305-3230	Telephone No. (703) 308-1096

for

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/05809

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,545,729 A (GOODCHILD et al.) 13 August 1996, see column 1-2, 4 and 9 and Figure 5.	20,34
Y	US 4,818,680 A (COLLINS et al.) 04 April, 1989, see abstract and Figure 2.	21
Y	US 5,407,795 A (KOLBERG et al.) 18 April 1995, see abstract and column 1.	22-24,29-33
Y,P	US 5,614,402 A (DAHLBERG et al.) 25 March 1997, see entire document.	1-68
Y,P	US 5,691,142 A (DAHLBERG et al.) 25 November 1997, see entire document.	1-68

THIS PAGE BLANK (USPTO)